Exploration and Discovery:
Learning at a Science Centre

Daniel Humberto Solís Recéndez

This thesis is presented for the degree of
Doctor of Philosophy
at the University of Otago
Centre for Science Communication

2019

Abstract

Free-choice learning provides the majority of human learning over the course of a lifetime. Since interactivity developed in science centres fifty years ago, hands-on learning has been spreading itself all around the world, with a current estimated visitation of 300 million people each year. This research investigated the impact of visiting a science centre on three aspects of scientific literacy: scientific knowledge, self-beliefs in science, and science engagement. While there is research around these constructs, the literature investigating the effect of a single visit on them using matched measurements before and after that visit is sparse. Data were collected with mixed methods, including pre-post surveys, interviews, and focus groups at the Otago Museum's science centre. The science centre is located in Dunedin, New Zealand, and recently underwent a major renovation. Before the redevelopment, 224 respondents filled out a pre-post survey. After the redevelopment, three more surveys were administered, collecting another 1,099 paired responses.

A brief 'formal' multiple-choice test proved to be useful in assessing knowledge without alienating visitors. A new self-reporting instrument was developed, called fluency in scientific concepts; results were closely related to scientific knowledge. Both constructs increased steeply with age from eight years old to early twenties and then changed only moderately. Also, both scientific knowledge and fluency increased significantly after one visit to the science centre, irrespective of age. Interviews with museum staff and two focus groups with children contributed to the research by identifying specific science exhibit characteristics that relate to visitor engagement and learning in all generations.

Self-efficacy and self-concept are related self-beliefs that behave differently. Self-reported self-efficacy increased dramatically with visiting the science centre, while self-concept remained stable. Although expected, these results have rarely been tested. Given its stability, self-concept was used to test a new alternative to Likert-type scales. A Visual Discrete Scale was developed for this research; it is characterized by being completely visual, with no labels. The Visual Discrete Scale proved to be a viable alternative to Likert-type scales with potentially more sensitivity to small changes in self-beliefs. A gender gap was detected in knowledge and self-beliefs. Evidence in this study did not support the idea of the gap being related to females' low-confidence. Instead, this study's results suggest that science engagement was originally low for female respondents, perhaps related to this study's focus on physics.

Dedication

To my parents, Emma Recéndez and Albino Solís. This is your triumph.

Acknowledgements

First of all, I wish to express my sincerest and upmost appreciaton to my knowledgable and always available supervisor, Professor Nancy Longnecker. Her day-to-day guidance and support were invaluable pillars in reaching this stage. I would also like to thank my co-supervisor, Professor David Hutchinson, who was always supportive in uncountable ways. I couldn't have found better supervisors—not only because of their academic achievements, but because of their extraordinary human qualities.

Over these years I also received feedback and advice from many people at the Centre for Science Communication. I'm particularly grateful to Andrea Liberatore, always willing to help. Also to Sue Harvey, who more than once helped me to survive bureaucracy. And to all the members of the Longnecker Research group for their invaluable time in reading and commenting on my methods.

Not being a native English speaker made the PhD especially challenging. I would like to acknowledge Rachel Leeson, who helped me to transcribe the interviews and understand the subtleties of meanings behind idioms. My gratitude also to Jean Fletcher, the first who made me discover the subtle variations in spoken English. Also to Jessica Davis, my personal editor who corrected my uncountable language-related mistakes.

Statistics is a cornerstone of this thesis. I want to acknowledge Professor Ben Daniel for his selfless sharing of expertise in the topic. I also want to thank my fellow PhD candidate and officemate, Ek Sripaoraya, for all the immensely helpful discussions we had about statistics.

Without the support of the Otago Museum, this research wouldn't have even started. My deepest thanks to Ian Griffin, Craig Grant, Sam Botting, Kate-Timms Deans, and Amadeo Enríquez-Ballestero. I also would like to thank all the science communicators and other staff for their support and friendship.

I would like to acknowledge all the visitors that accepted participating in the surveys. Also to the children who donated part of their sleepover time to participate in the project. Especially, I would like to infinitely thank the interviewed museum staff and the focus groups participants; their insights were invaluable for understanding science engagement and learning at the science centre.

Without financial support, I wouldn't be at this life-changing stage. I will always be grateful to the University of Otago for granting me the Doctoral Scholarship. Also to the Dodd-Walls Centre for Photonic and Quantum Technologies for the multiple times they backed the research. All my gratitude also to the Instituto Politécnico Nacional and the

Unidad Profesional Interdisciplinaria de Ingeniería Campus Zacatecas for their support through the Licencias con Goce de Sueldo CPE/COTEBAL/105/2016, CPE/COTEBAL/100/2017, CPE/COTEBAL/71/2018 and CPE/COTEBAL/67/2019.

Moving to a new country with a different culture and language, and not knowing anyone is daunting and can affect performance. All my appreciation to the people that helped me face it. My hearthfelt thanks to my family and friends. Especially to Emma Solís, to whom I constantly asked for advise about my always shape-shifting thesis, and to Víctor Nuñez and Eduardo Lara, the never fading musketeers.

To all the people that helped me in this journey in one way or another, many thanks.

Table of Contents

List of Tables

List of Figures

List of Appendices

Images Authoring

Some of the images used in the thesis have copyright, but can be used for free in non-commercial environments. Copyrighted authoring is acknowledged in the following table:

Identifying name	**Author**	**URL**
Puzzle	Kraphix	freepik.com/free-vector/business-cartoon-with-puzzle_1013675.htm
Children around the world	Rosapuchalt	freepik.com/free-vector/sketchy-kids-around-world_783156.htm
Penguin in suit / Seal	GraphicMama-team	pixabay.com/illustrations/polar-animals-antarctic-arctic-1453993/
Eyes	Freepik	freepik.com/free-vector/comic-style-eyes_761062.htm
Jumping penguin	OpenClipart-Vectors	pixabay.com/vectors/penguin-blue-little-animal-156616/
SWOT theme	Freepik	freepik.com/free-vector/swot-analysis-themes_716498.htm
Stop	TheDigitalArtist	pixabay.com/illustrations/stop-process-business-solution-1829082/
Gender icons	Smalllikeart	flaticon.com/free-icon/toilet_895375
Painter penguin	OpenClipart-Vectors	pixabay.com/vectors/painter-tux-art-artist-beret-161318/
Owls	KirstenStar	pixabay.com/vectors/owl-owls-cartoon-animal-nature-1814023/
Painting palette and brush	Pch.vector	freepik.com/free-vector/art-studio-illustration_4979636.htm

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

- 1.1 Introduction
- 1.2 Why Science Communication Matters
- 1.3 Conceptual Frameworks
- 1.4 Learning at Science Centres
- 1.5 Assessment of Learning: the gaps
- 1.6 Uncovering Learning
- 1.7 Summary

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 1: INFORMAL LEARNING OF SCIENCE

1.1 Introduction

This thesis aims to uncover aspects of how science learning happens in science centres. This chapter introduces why communicating science is important in a knowledge society. It includes a conceptual framework where “science learning” is defined by the increase of factual knowledge, engagement, and self-beliefs. In this framework, science communication is a key element in the process of public science learning. Previous research about science learning in science centres, and the role engagement plays, is presented and discussed before introducing the key research questions addressed by this thesis.

1.2 Why Science Communication Matters

1.2.1 The hydra of bad science

In the era of the Internet, an amazing amount of information is easily published and freely and immediately available (Kilgarriff, 2007). For example, from 2009 to 2010, the number of videos on YouTube doubled, with approximately 24 hours of content uploaded each minute (Patiño, 2013). Fake news (false information presented as true) is available via this unlimited publishing, posing a serious threat to real information (Lazer et al., 2018; Lockwood, 2010; Shao et al., 2018; Shu, Sliva, Wang, Tang, & Liu, 2017).

Currently, the world is being challenged by global threats like climate change. Only by acting together can we do something about it. But bad science is a hydra that grows two new heads every time a scientist cuts one off. This handicap is due to the meticulous work science requires, compared to the ease with which other information is produced.

1.2.2 The teamwork of Heracles and Iolaus

In Greek mythology, when Heracles was assigned the task of killing the Hydra, he asked his friend and nephew, Iolaus, to work with him. Each time Heracles cut off one of the monster’s heads, Ioulaus would seal the wound with a hot iron to avoid new heads being born. Killing the Hydra of Bad Science is destined to fail if one party acts alone. The different roles Heracles and Iolaus played in this analogy are not as important as the message: we all

–scientists, science communicators, and citizens– need to work together to prevent the Hydra from growing more heads.

The task may seem gargantuan, but it is feasible. Populations around the world have generally stable attitudes towards science (National Academies of Sciences, Engineering, and Medicine, 2016). “[A] thoroughly positive picture of young people’s appreciation of science emerges worldwide” (Krapp & Prenzel, 2011, p. 41). Science centres can directly contribute to public understanding of science (Falk & Needham, 2011). Although citizens do not necessarily need to use their understanding of science in daily life, scientific literacy is useful for participating in a society driven by technological advancements (Feinstein, 2011). Scientific progress may depend on the public understanding science (Pasek, 2018), as science related policies are, at least in part, based on public perceptions (Gauchat, 2011).

1.2.3 The seed of science communication

Informal learning is driven by curiosity (Patiño, 2013), and interest in science usually starts before middle school (Maltese & Tai, 2010). It is during primary and lower secondary school when informal experiences with science have their strongest influence (Venville, Rennie, Hanbury, & Longnecker, 2013). It is estimated that half of the visitors to science and discovery centres in England are children (Ecsite-uk, 2008). Dr. Griffin, director of the Otago Museum, is an example of what can happen when you communicate science to young people. His parents couldn’t afford to pay to go to the science museum in London, but when he was four or five years old they took him on a free day.

> I remember to this day going through and turning those handles and seeing these machines doing amazing things. And I put down that visit to inspire me in seeking a deeper interest in science … Museums are a cog in the wheel that can potentially create a new generation of passionate citizens. (Ian Griffin, personal communication, 20 September, 2018).

That visit pushed Dr. Griffin to later become an astronomer, and eventually a museum director. “I know that museums have the potential to transform people’s lives”, he added. He is a living example of that transformation.

Activities outside the classroom influence students’ self-efficacy and perceptions of STEM careers (Mohtar et al., 2019). Learning outside the classroom improves academic achievement, makes learning more engaging and relevant, nurtures creativity, and improves young people’s attitudes toward learning (Ecsite-uk, 2008).

1.3 Conceptual Frameworks

1.3.1 Science communication

The investigation of informal science learning rests upon how exactly science is communicated. The widely used deficit model assumes that the public has negative attitudes towards science because of ignorance. To solve this ‘deficit’, information flows in one direction from experts (knowledge holders) to the public (knowledge novices). But this method of communication is unsuccessful within polarizing topics like climate change. It is well-established that some people reject anthropogenic factors in climate change even when they are made aware of the scientific facts (Kahan et al., 2012). Indeed, they hold onto personal beliefs that contradict the scientific consensus (Roos, 2014). Thus, it may not be ignorance, but motivated reasoning, that explains an individual’s rejection of scientific information (Pasek, 2018).

The assumption that exposing people to more science will cause them to embrace and support it does not hold up (Allum, Sturgis, Tabourazi, & Brunton-Smith, 2008; Phillips & Beddoes, 2013). More knowledgeable individuals are more consistent in their defence of their beliefs (Drummond & Fischhoff, 2017; Kahan et al., 2012). It is not knowledge alone that determines how some will react, but a person’s values, beliefs, attitudes (Ahteensuu, 2012; Buddle, Bray, & Ankeny, 2018; Cortassa, 2016; Hart & Nisbet, 2012; Kahan et al., 2012), affect (Carver, 2001), existing cognition (Bucchi, 2008; Buddle et al., 2018; Kahan et al., 2012), personal meaning of science (Bucchi, 2008; Falk & Dierking, 2016), and personal relevance (Falk & Dierking, 2016; Ham, 2016; Harré, 2011). Moreover, when a narrative is accepted and incorporated into personal understanding, it becomes hard to change (Cook & Lewandowsky, 2016; Larson, Cooper, Eskola, Katz, & Ratzan, 2011), especially if that understanding is shared by the person’s social network (Harré, 2011).

An alternative to the deficit model is the Koru Model of Science Communication (Longnecker, 2016). This model states that evidence alone is not enough to convince people to change their beliefs or practices, and uses the analogy of a koru—a fern that symbolizes life and growth in New Zealand Māori culture. There are three main components in this model: communication, engagement, and use of information. Informal education would be part of the roots (channels of communication) that help to absorb the facts (nutrients) and transform them into coherent information that is transmitted to the organism or vice versa. In response, the individual can engage with the information if this information fits with existing knowledge, values, and attitudes. The Koru Model lets us visualize how the flow and use of information is determined by a recipient's identity. The Koru Model (Figure 1-1) guides this research.

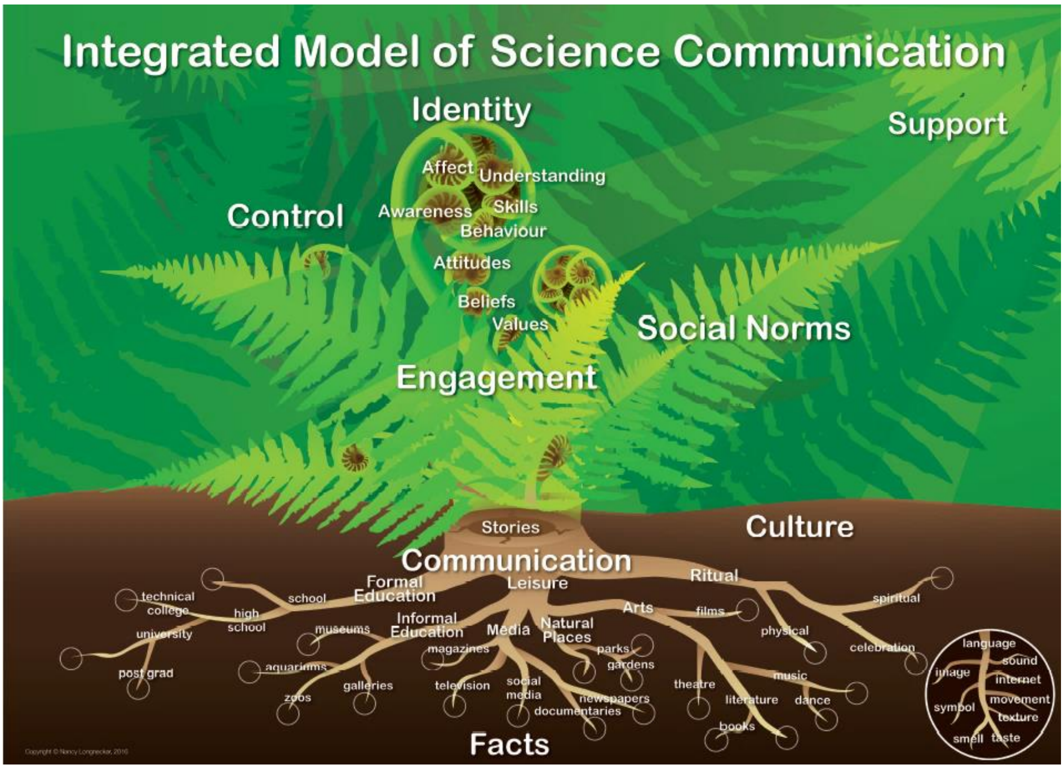


Figure 1-1. Detailed koru integrated model of science communication. Reprinted from Longnecker (2016) with permission of the author.

1.3.2 Science learning

Learning is a familiar concept, but due to its complexity, no unanimous definition exists (Barron et al., 2015). It is traditionally defined in terms of knowledge acquisition (Fender & Crowley, 2007; Illeris, 2018). But learning comprises more than knowledge gains. It includes changes in understanding, feelings, and attitudes (Illeris, 2018; Krishnamurthi & Rennie, 2012), self-related cognitions, interests, expectations, behaviours, and life-skills (Falk, 2005; Organisation for Economic Cooperation and Development [OECD], 2009). Learning has been increasingly defined in terms of behavioural change (Barron et al., 2015; De Houwer, 2011; De Houwer, Barnes-Holmes, & Moors, 2013; Pear, 2014).

Learning is "a structured updating of system properties based on the processing of new information" (Barron et al., 2015, p. 406). But to define what system properties can be used to assess that learning, four complementary frameworks were considered. The first framework, Constructivism, is an epistemology of the nature of learning. Its central idea is that human learning is actively constructed, built on prior personal knowledge (Cobern, 1993; Simpson, 2002). Learning implies creating a cognitive conflict in students' minds—a challenge that confronts their prior conceptions with new phenomena and new knowledge, producing amendments to their understanding (Bächtold, 2013; Cobern, 1993; Simpson, 2002). Grigorovitch (2014) found that a constructivist strategy in class is more effective for children to understand the concept of light, rather than a classical teaching approach. The hands-on nature of science centre activities provide a great environment for learning under the constructivism theory (Eshach, 2007)

The second framework is Situated Learning. It considers that learning is favoured by active participation in a social environment (physical and social contextualization) connecting with prior knowledge in an authentic, informal context (Gee, 2012; Lave, 1988; Lave & Wenger, 1991; Mattar, 2018). Conceptual understanding is constructed by interacting with objects and phenomena (Renner, 2011). Notwithstanding, it is important to acknowledge that abstraction instruction (not situated) can also be very effective in some cases (J. R. Anderson, Reder, & Simon, 1996).

Most people identify interest and curiosity as their main motivations to acquire knowledge (Falk, Storksdieck, & Dierking, 2007; Feher, 1990; Venville et al., 2013). But motivations are shaped by people's specific needs, abilities and socio-historical context (Falk, Storksdieck, et al., 2007). The third framework is the Contextual Model of Learning (Falk & Dierking, 2004, 2016, 2018; Falk & Storksdieck, 2005). This model considers that visitors decide what to focus on based on their personal context (e.g., personal motivations), physical context[1] (e.g., benches to sit) and social context (e.g., a visit for an adult is different if accompanied by peers than by children).

The fourth framework is the Koru Model (Longnecker, 2016). The Koru Model integrates aspects from the other frameworks. It utilises constructivism by assuming new information builds upon an existing framework. Use of new information is situated; it is fostered by support, influenced by social norms, and ultimately affected by whether the individual has control over their ability to use the information.

In summary, learning is constructed by challenging previous understanding. It is affected by personal and external contexts. And it favours active participation—not only personal, but social. Moreover, everybody can learn (McCombs, 2001).

[1] 'Physical context' (contextual model of learning) and 'physically contextualized' (situated learning) are not the same. The former refers to the effect of the physical space on the experience, like lighting, places to sit, etc. The latter refers to the relation between the setting and what is learnt, e.g., learning to swim in a pool as opposed to a classroom.

1.3.3 Beyond the facts: three branches of scientific literacy

There have been many definitions of scientific literacy throughout the history of the term. Some have been intellectual, others attitudinal, societal, or interdisciplinary (Holbrook & Rannikmae, 2009). There is no consensus on its definition (DeBoer, 2000). "[T]he only way to avoid confusion about SL [scientific literacy] is to stipulate its meaning every time one uses the term" (Linder et al., 2010, p. 13).

Even when scientific information reaches the audience, it may not be integrated into the audience's framework if the scientific evidence does not fit with existing knowledge and cultural philosophies or values (Kahan et al., 2012; Longnecker, 2016). Individuals may just reject evidence that contradicts their previous understanding of the world (Lodge & Taber, 2013), or how it is integrated may be biased (National Academies of Sciences Engineering and Medicine, 2018). Kahan et al. (2012) studied perceptions about climate change in 1,540 US adults, which depended on their level of science literacy and numeracy, as well as how egalitarian/communitarian and hierarchical/individualistic their worldviews are. Their results indicate that public rejection of climate change does not come from lack of science understanding, but from a conflict of interest between what science says and what their own interests are, including those of the community they interact with. For example, among hierarchical individualists, polarization increases as science literacy and numeracy do. Scientific knowledge can cause polarization of beliefs depending on whether or not facts fit into people's worldviews.

The OECD Programme for International Student Assessment (PISA) Framework considers scientific literacy more comprehensively. It involves not only knowledge and attitudes (and values and beliefs), but the use of that knowledge and engagement with science-related issues (OECD, 2006, 2016b). PISA assesses scientific literacy internationally in 15 year-old students in three main subjects: Science, Mathematics, and Reading. Under the PISA Framework for scientific literacy (OECD, 2006, 2010, 2013a, 2016a, 2016b), science was the main subject in 2006 and 2015 (OECD, 2010, 2016b; Stacey, 2010).

Scientific literacy is applied in personal, social, and global situations (OECD, 2014). It can be assessed in relation to competencies such as scientific knowledge or concepts (identification), scientific processes (explanation), and situations or contexts (use). These competencies require not only knowledge and cognitive abilities, but attitudes, values, and motivations (OECD, 2006, 2010, 2013a) to be able to make more informed choices[2] (OECD,

[2] Note how this framework coincides with the explanation of the Koru Model to why people can know facts about climate change and still deny it.

2016b). In a Knowledge Society, the focus is less on what you know and more on what you can do and learn with that knowledge (Linder et al., 2010). Engagement, interest, motivation, and belief in self-capacities foster learning (Christenson, Reschly, & Wylie, 2012; OECD, 2013b, 2016b; Schunk & Mullen, 2013).

It is not the aim of this research to uncover every possible effect of visiting a science centre. "Even the most broadly defined learning outcomes may not be sufficient to explain the value and benefits of the museum experience" (Packer, 2008, p. 33). To operationalize scientific literacy as a measurable outcome, several frameworks were considered (Table 1-1). All of the frameworks included Knowledge and Engagement as important outcomes. However, the framework should also consider self-beliefs.

Table 1-1
Summary of frameworks that include in their impacts knowledge (K), self-beliefs (SB) or engagement (E)

Framework	**Reference**	**K**	**SB**	**E**
AEIOU Framework	T. W. Burns, O'Connor, and Stocklmayer (2003)	●	×	●
Framework for Evaluating Impacts of Informal Science Educations Projects	A. J. Friedman (2008)	●	○	●
Generic Learning Outcomes Framework	Hooper-Geeenhill, Dodd, Morrison, and Toon (2003)	●	●	●
Strands of Science Learning Framework	National Research Council (2009)	●	●	●

The closest outcome in the AEIOU framework (T. W. Burns et al., 2003) to Self-Beliefs is Opinions, but it does not include beliefs about one's self.Attitude (i.e. towards STEM-related topics or capabilities) is the closest category to self-beliefs in the Framework for Evaluating Impacts of Informal Science Educations Projects (A. J. Friedman, 2008), however its inclusion of confidence is tangential. The Generic Learning Outcomes (Research Centre for Museums and Galleries, 2003; Hooper-Greenhill, 2002) considers change in Values, attitudes, and feelings as an outcome. The importance of attitudes towards self (self-esteem and confidence, among others) in this framework is clearly stated.

The Strands of Science Learning Framework (National Research Council, 2009) was developed based on the Strands of Scientific Proficiency Framework (National Research Council, 2007). It includes six strands; the sixth "addresses how learners view themselves with respect to science" (National Research Council, 2009, p. 4), which fits with how self-beliefs are defined in this work. This framework is also presented by Fenichel and

Schweingruber (2010), but more succinctly—focusing on real life examples, rather than the theoretical background. The Strands of Science Learning Framework considers that "science is a system of acquiring knowledge through systematic observation and experimentation" (National Research Council, 2009, p. 42). Three strands (outcomes) in this framework are relevant for this particular research:

- Strand 1: Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world.
- Strand 2: Come to generate, understand, remember, and use concepts, explanations, arguments, models, and facts related to science.
- Strand 6: Think about themselves as science learners and develop an identity as someone who knows about, uses, and sometimes contributes to science.

Strand 1 will be referred in the rest of the thesis as "science engagement", Strand 2 as "scientific knowledge", and Strand 6 as "self-beliefs in science". Note that science engagement, scientific knowledge, and self-beliefs in science are interrelated and they are referred to by different names in the literature. For example, since scientific knowledge is synergistically connected to comprehension (Cervetti & Hiebert, 2015), it is sometimes used as a synonym of scientific literacy (e.g. Kahan et al., 2012; National Science Board, 2018). Self-efficacy can also be considered part of engagement (OECD, 2006). The term Science Fluency, a self-belief in science later in this chapter, is sometimes used to refer to something similar to scientific knowledge (Ceuppens, Deprez, Dehaene, & De Cock, 2018; Hill & Sharma, 2015; Hill, Sharma, & Johnston, 2015; Hill, Sharma, O'Byrne, & Airey, 2014; McCallie et al., 2009; Patel, 2013; Powers & Kier, 2016; Wieland, 2015).

To avoid mixing terminologies, throughout this thesis terms associated with scientific literacy will follow the definitions given in Table 1-2, regardless of the name with which the construct is used elsewhere.

Table 1-2
Definitions of the principal terms related to scientific literacy

Term	Definition	Reference
Scientific literacy	Understanding of natural phenomena and how they influence the world and society. Includes self-awareness of possessing such understanding, and willingness to engage with science-related issues required to make decisions as a reflective citizen.	Author, based on OECD (2006)
Scientific knowledge	Knowledge of facts, concepts, ideas, and theories about the natural world that science has established.	OECD (2016b, p. 16)
Self-beliefs in science	Perceptions of oneself with respect to science, that allow a person to develop an identity as someone who knows and can learn more science.	Author, based on National Research Council (2009)
Science engagement	Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world.[3]	National Research Council (2009, p. 4)
Science	System of acquiring knowledge through systematic observation and experimentation.	National Research Council (2009, p. 42)
Science communication	Use of appropriate skills, media, activities, and dialogue to produce a personal change in scientific literacy when exposed to science.	Author, based on T. W. Burns et al. (2003)
Science learning	Structured updating of scientific literacy based on processing new information that challenges a prior state.	Author, based on Barron et al. (2015)

Scientific knowledge

Scientific knowledge is a vital component of scientific literacy (American Association for the Advancement of Science, 1994; National Academies of Sciences, Engineering, and Medicine, 2016). There are different types of scientific knowledge. The one considered in this thesis is content knowledge, as defined in Table 1-2. Though limited in scope, content and procedural knowledge are reasonable indicators of science knowledge (National Academies of Sciences, Engineering, and Medicine, 2016).

The lesser the prior knowledge, the greater the gain can be (Falk & Storksdieck, 2005). But since previous knowledge helps in making sense of the world (Arts Council England, 2019), visitors are more attracted to exhibits that relate to something they already have some knowledge of (Falk & Dierking, 2016). But it is important to note that each individual has his or her own science knowledge repertoire and science understanding (Falk, Storksdieck, et al., 2007).

[3] This definition is not a definition of science engagement per se, it is the description of Strand 1, which fits with our understanding of science engagement.

Following the constructivist theory of learning, "All learning involves transfer from previous experiences. Even initial learning involves transfer that is based on previous experiences and prior knowledge" (National Research Council, 2000, p. 68). Prior knowledge is almost an unbounded variable in informal settings, as visitors range from novices to experts (Schwan, Grajal, & Lewalter, 2014).

Whether or not that previous knowledge is 'right' is a different matter. Prior knowledge greatly affects the learning experience; it can both facilitate and interfere (Roschelle, 1995). Stocklmayer and Bryant (2012) compared results from a survey of 4,000 members of the public and 500 scientists. Scientists were allowed to discuss their wrong answers and two remarkable things happened. First, no one felt completely confident of their answers to every question, with many admitting ignorance outside their field. Second, some scientists justified wrong responses in interesting ways. For example, some scientists selected false to the question on whether hot air flows up, and justified they got it wrong by saying that, technically, it is cold air what goes down, or that hot air rising is only valid if a gravitational field is present. It is then perfectly plausible that wrong answers by the public do not mean ignorance, but a different logic.

Also, knowing something is not binary; it is more of a continuum where someone may not know details, but has 'some' knowledge (J. K. Smith, 2014). Or, it can be that a person is able to identify the answer in a different format, such as visually (Bucchi & Saracino, 2016).

Self-beliefs in science

Confidence is important in science; people are more likely to try something if they feel they can be successful (Fenichel & Schweingruber, 2010). Three self-beliefs in science are relevant to the research described in this thesis: self-concept in science, self-efficacy in science, and scientific fluency. An explanation of each and the reasons for their selections are provided below. Table 1-3 presents a summary of definitions of the terms related to self-beliefs in science.

Table 1-3
Definitions of terms related to self-beliefs in science

Term	Definition	Reference
Self-concept in science	Individual's general perception of their own abilities in science.	Author, based on OECD (2018)
Self-efficacy in science	Perceived capacity for doing specific science-related tasks in given situations.	Author, based on OECD (2018)
Scientific fluency	Perceived science-related knowledge that facilitates comprehending natural world phenomena.	Author, based on J. K. Smith (2014)
Scientific familiarity	Perceived knowledge about specific science term or scientist.	Author, based on J. K. Smith (2014)

Self-concept is an individual's general perception of their own abilities related to doing well in a given domain (Bong & Skaalvik, 2003; Huang, 2011; Jansen, Scherer, & Schroeders, 2015; OECD, 2018; Wilkins, 2004). It is fairly stable (Lee, 1998), but it is heavily influenced by social comparison (Bong & Skaalvik, 2003; Jansen, Schroeders, & Lüdtke, 2014). An individual's self-concept is affected by previous performances in the area and comparisons with performances in other areas (Jansen et al., 2015). It mutually reinforces achievement (DeBacker & Nelson, 2000; Huang, 2011; Jansen et al., 2015; Jansen et al., 2014; Marsh & Martin, 2011; Marsh, Xu, & Martin, 2012; Wender, 2004; Wilkins, 2004). Self-concept in science is an important component of self-beliefs in science (OECD, 2009; Wilkins, 2000, 2004), and it can influence career aspirations in science (Nagengast et al., 2011). Actually, it can influence attitudes and behaviour in general, "most humans, most of the time, tend to *act in accordance with the image that they have of themselves* [emphasis in original]" (R. S. Miles, Alt, Gosling, Lewis, & Tout, 1988, p. 25)

Self-efficacy is an individual's perception of their own capacity for doing specific tasks in given situations (Bandura, 1986; Bong & Skaalvik, 2003; OECD, 2009, 2018). It is also strongly affected by personal experiences (Jansen et al., 2015), but it is less influenced by relative impressions (Bong & Skaalvik, 2003; Jansen et al., 2015). Self-efficacy is also strongly linked to achievement (Bandura, 1986; Bandura & Locke, 2003; Diseth, Meland, & Breidablik, 2014; Honicke & Broadbent, 2016; Komarraju & Nadler, 2013; Valentine, DuBois, & Cooper, 2004; Zajacova, Lynch, & Espenshade, 2005; Zeldin, Britner, & Pajares, 2008).

Fluency refers to the ease with which information flows through the cognitive system, such that the individual is able to extract information with lower effort and attentional load (Kellman, Massey, & Son, 2010; Shimamura & Palmer, 2012; J. K. Smith, 2014; L. F. Smith

& Smith, 2006; Tsai & Thomas, 2011). Fluency is related with the familiarity an individual has with something (J. K. Smith, 2014)[4].

An example of an instrument similar to Science Fluency was developed by Emereole (2009), who asked university science students and senior secondary school science teachers to rate their level of familiarity (Not familiar / Uncertain / Very familiar) with 15 science processes (e.g., Observation, Controlling variables). Mbewe, Chabalengula, and Mumba (2010) modified Emereole's instrument to study pre-service teachers' familiarity with science process skills. The new instrument had 13 science processes skills (e.g. Classification, Controlling variables) with three responses (Term not familiar to me / Term familiar to me but I do not understand its meaning / Term familiar to me and I understand its meaning).

The instruments above can be adapted to the museum setting, but there is a third option coming from the art realm. *Aesthetic fluency* is the knowledge a person has regarding art (J. K. Smith, 2014; L. F. Smith & Smith, 2006). The instrument developed by Smith and Smith (2006) to assess aesthetic fluency asks individuals how much they know about five artists and five art ideas. Respondents have five options to choose from:

- I have never heard of this artist or term
- I have heard of this but don't really know anything about it
- I have a vague idea of what this is
- I understand this artist or idea when it is discussed
- I can talk intelligently about this artist or idea in art

Smith and Smith's instrument was deemed more suitable for this research for several reasons. It was used with museum visitors. Assessing fuency about artists and art terms can easily be transformed into assessing scientists and science terms. There are five options, which allows for better discrimination than having three. Lastly, the model appears in more research than the alternatives (see Atari, Afhami, & Mohammadi-Zarghan, 2018; Fayn, Silvia, Erbas, Tiliopoulos, & Kuppens, 2018; Silvia, 2007, 2013; L. F. Smith & Smith, 2006).

It is not strange that an instrument designed to measure art fluency is so well suited to measure Science Fluency. Fluency has aesthetic qualities. People tend to prefer and positively value pictures they recognise; i.e., we like what we feel we are fluent in (Belke, Leder, Strobach, & Carbon, 2010; Forster, Leder, & Ansorge, 2013; Shimamura & Palmer,

[4] Do not confuse with fluency as the ability to read with sufficient ease and accuracy. Ganeb and Morales (2018) call 'science fluency' the ability to read scientific text and understand its meaning, but this is not how this research defines it.

2012; J. K. Smith, 2014; Tsai & Thomas, 2011). But aesthetic is also an epistemic value[5] in science; "aesthetic experiences can function as normative judgements, and thus have normative consequences regarding what should be included and excluded in the learning process" (Linder et al., 2010, p. 164). How aesthetic appreciation impacts learning goes beyond the aim of this work.

According to L. F. Smith and Smith (2006) and J. K. Smith (2014), aesthetic fluency is the knowledge a person has regarding art. It is "understanding what *chiaroscuro*[6] means *and* being able to spot its use in a painting [emphasis in original]"[7] (L. F. Smith & Smith, 2006, p. 6). Therefore, the instrument developed by Smith and Smith does not measure aesthetic fluency in its full extent. It relies on self-reports and does not measure if a respondent can effectively 'spot its use in a painting'.

In a study about the relation between factual knowledge and familiarity, Ladwig, Dalrymple, Brossard, Scheufele, and Corley (2012) measured factual nanotechnology knowledge with six true/false questions and perceived nanotechnology familiarity with a 10-point scale ('How well informed you would say you are about nanotechnology?'). Pearson's correlation between factual knowledge and familiarity, although significant ($p<.001$), turned out to be small ($R=.187$). They conclude that factual knowledge and perceived familiarity refer to two distinct dimensions of understanding and should be considered conceptually distinct.

Therefore, the research reported in this thesis limits what the scientific fluency instrument claims to measure. It is a self-reported self-belief, rather than an empirically and objectively measured construct. Scientific fluency is here defined as the *perceived* science-related knowledge that facilitates comprehending natural world phenomena (see Table 1-2).

Science engagement

Engagement is a multidimensional construct that includes behavioural, emotional, social, and cognitive engagements (Olitsky & Milne, 2012; M.-T. Wang, Fredricks, Ye, Hofkens, & Linn, 2016). It is an important aspect of scientific literacy (McCallie et al., 2009; OECD, 2006). Engagement creates the opportunity for meaningful and positively felt experiences (National Research Council, 2007). It is considered a stepping stone to science learning (Ainley & Ainley, 2011; McCallie et al., 2009).

---

[5] Crucial values for a specific activity.
[6] Chiaroscuro is a painting technique where light and shadow are strongly contrasted.
[7] Emphasis comes from source.

Enjoyment and interest are closely associated to engagement (Ainley & Ainley, 2011). People interested in science are more likely to be motivated science learners who use strategies for effective learning and continue to get further engaged (Csikszentmihalyi, Rathunde, & Whalen, 1997; Renninger & Hidi, 2002). Learning in informal environments is a function of personal motivations, interests and opportunities (C.-C. Liu & Falk, 2014; National Academies of Sciences Engineering and Medicine, 2018). Interest may not only lead to higher academic achievement, but a pursuit of science-related careers (Krapp & Prenzel, 2011).

1.3.4 Moderating variables

A moderating variable[8] (a.k.a. moderator) is an exogenous variable that affects the relationship between the independent and dependent variable by amplifying, diminishing or qualitatively altering the influence one has on the other (Baron & Kenny, 1986; Dearing & Hamilton, 2006; A. F. Hayes, 2017). Two common moderating variables are age and gender (Smyth, 1998; Wehmeyer et al., 2011). However, it is important to keep in mind that differences are averages, no predictions can be done from group data to individuals. For example, "within each sex there is almost the full possible range of abilities or types of behaviour" (Lindon, 1996, p. 22).

Age

How people learn varies with age (Fenichel & Schweingruber, 2010). Knowledge also grows with age, especially in young people (Lindon, 1996). One important transition in mental maturity is puberty, and pubertal changes do not occur at the same time in girls and boys (Eccles, Templeton, Barber, & Stone, 2003). However, age is an attribute where not only biology plays a role, but also socialization, experience and cultural expectations (Laz, 1998). Moderating variables in single constructs can have a normative age-graded influence, i.e., correlate with chronological age for most respondents (Baltes & Nesselroade, 1984; Plomin, 1986).

Piaget's theory of cognitive development (Piaget, 1968) holds that children's thinking, by incorporating inductive reasoning, becomes more logical and organized during what he calls the Concrete Stage of Cognitive Development. This starts at around seven to eight years old. Around 11-12 years, in the Formal Operational Stage, a person's logic and

[8] Do not confuse it with a mediator variable, which does not affect the strength of the relationship, but explains how and why the intervention is producing those effects. However, notice that Creswell (2009) defines the mediator in the same way the moderator is defined here, and gives the moderator a quite different definition.

ability to use deductive reasoning increases, and they become better at understanding abstract ideas.

In agreement with the theory of cognitive development, Lindon (1996) explains that children from eight to 11 years start understanding symbology and gain experience mainly from inductive reasoning, from particular experiences to the general principle. From 12 or 13 years of age is when children are able to handle deductive logic and abstract ideas. Piaget’s theory of cognitive development does not include a stage that separates adolescents from adults. But he considers more complete logic in the Formal Operational Stage is not reached until 14 to 15 years old, with the possibility of extending up to 20 years old, when professional specialization begins (Piaget, 1972).

The relationship between the brain and learning is reciprocal. Learning organizes and reorganizes the brain, even at the neural connections level (National Academies of Sciences Engineering and Medicine, 2018; National Research Council, 2000). Emotion regulation also has important implications in cognitive development (Silvers et al., 2012). But even when the brains of adolescents aged 13 to 17 years have neurological similarities with adult brains, they have not yet reached adult levels of working memory and internal control (Crone & Ridderinkhof, 2011).

These indicate that another division could be added to Piaget’s theory—one that separates adolescents from adults. But when an adolescent becomes an adult varies depending on the perspective. To some, adulthood starts at 16 (Borgers, Hox, & Sikkel, 2004), 17 (Silvers et al., 2012), 19 (Lesko, 2012), or even up to 24 or 25 years old (Jančić, 2016; Sawyer, Azzopardi, Wickremarathne, & Patton, 2018). This will be discussed further in Chapter 3.

Gender

In 37 of 54 countries and economies that participated in PISA 2006 and 2012, no gender gap in science was detected. In fact, in ten countries girls outscored boys, mainly in identifying scientific issues (OECD, 2014). However, other reports show a gender gap where females score lower than males in scientific literacy (S. Allen, 1997; Kurtz-Costes, Rowley, Harris-Britt, & Woods, 2008; Skaalvik & Skaalvik, 2004) and self-concept in science (Jansen et al., 2015; Kurtz-Costes et al., 2008).

Kanyangarara, Mayberry, Pai, and Shanahan (2012) reviewed the most commonly used gender frameworks (Harvard Analytical Framework, Women’s Empowerment Framework, Moser Framework, Social Relations Approach, and Gender Analysis Matrix).

None fit with the approach of this research. However, no specific model needs to be considered. What matters is acknowledging when students hold gender stereotyped views of science, they are less likely to pursue learning goals in science (DeBacker & Nelson, 2000). Low self-esteem in science (Bamberger, 2014), discrimination, stereotypes, restrictive curricula, lack of encouragement, and lack of opportunities still affect female achievement (Reilly, 2010; UNICEF, 2007; Vimala, 2010). Stereotypes associated with males' high-level intellectual ability influence children as young as six and can discourage women from pursuing careers such as physics (Bian, Leslie, & Cimpian, 2017).

1.3.5 An integrated framework

The approach of this research is pragmatic. By not committing to a single philosophical system, all approaches are available to understand the problem (Creswell, 2009). Figure 1-2 represents a framework that integrates the frameworks discussed above and underpins the research described in this thesis. It does not address all the possible factors influencing science learning, but it includes the main ones that are considered in the research. Learning is seen as a change in scientific literacy (in any of its three components: scientific knowledge, science engagement, and self-beliefs in science). Science Communication is the palette that helps to increase scientific literacy. The individual's age, gender, contexts, and prior knowledge form the structure that supports learning.[9]

[9] Image created by editing and modifying the image *art studio illustration* by pch.vector from freepik.com.

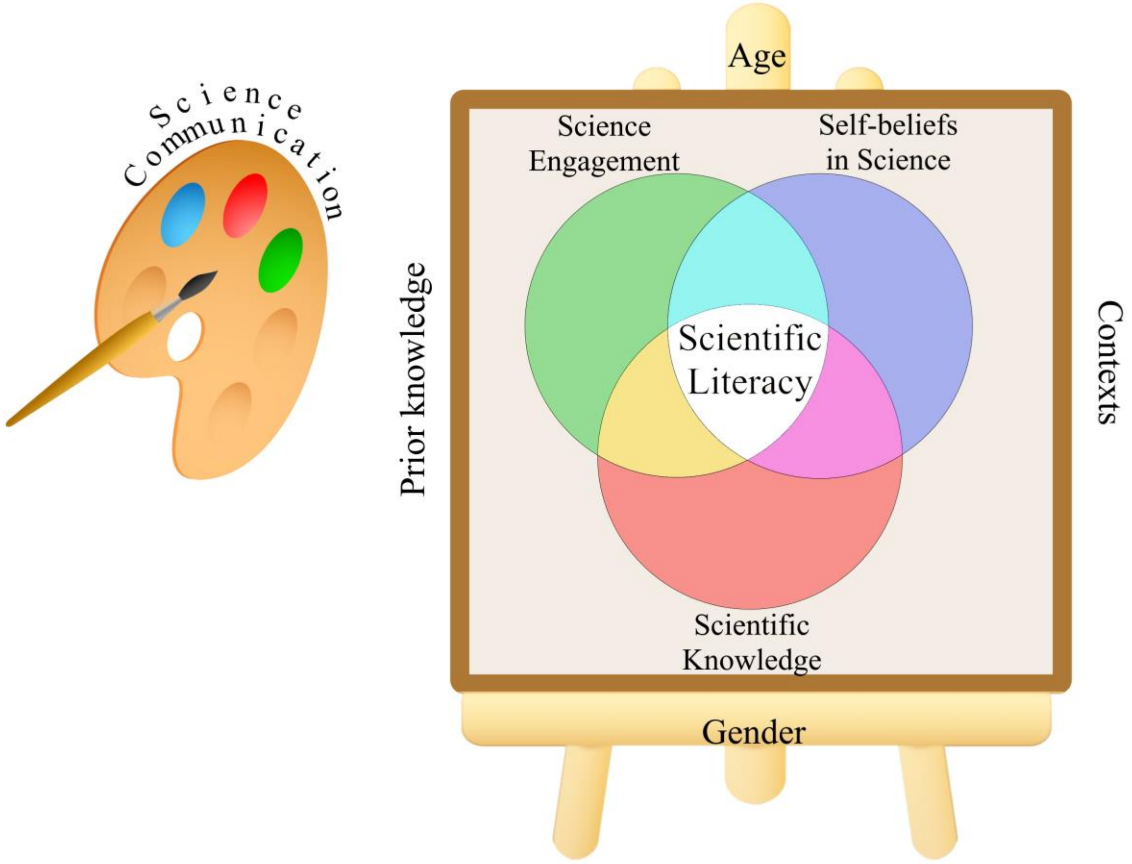


Figure 1-2. Integrated framework of science learning used in development of research questions for this thesis.

In this framework, learning is understood as a change in scientific literacy. Scientific literacy comprises scientific knowledge, science engagement, and self-beliefs in science. Science communication (painter's palette) acts as an intermediary to help an individual increase their scientific literacy. Which of these components increases the most depends on the individual's age, gender, contexts[10], and prior knowledge.

1.4 Learning at Science Centres

1.4.1 Free-choice informal learning

There are three main forms of learning, according to the taxonomy of education: formal, nonformal, and informal (La Belle, 1982). Based on a comprehensive description by Werquin (2007), Ainsworth and Eaton (2010) summarized their characteristics (Table 1-4).

[10] Physical (e.g. design of exhibits), personal (e.g. visit motivation) and sociocultural (e.g. cultural background).

Table 1-4
Definitions of formal, nonformal and informal learning

Formal learning	This type of learning is intentional, organized and structured. Formal learning opportunities are usually arranged by institutions. Often this type of learning is guided by a curriculum or other type of formal program.
Nonformal learning	This type of learning may or may not be intentional or arranged by an institution, but is usually organized in some way, even if it is loosely organized. There are no formal credits granted in nonformal learning situations.
Informal learning	This type of learning is never organized. Rather than being guided by a rigid curriculum, it is often thought of as experiential and spontaneous.

Reproduced with permission from Ainsworth and Eaton (2010).

Notice how the three definitions above are independent of personal intention. For instance, an afterschool program is nonformal learning, but may be not be voluntary if parents enrol their children and expect them to attend (Krishnamurthi & Rennie, 2012). It is also important to clarify that even when learning can be divided according to its formality, almost all learning situations contain attributes of both informality and formality (Malcolm, Hodkinson, & Colley, 2003).

Free-choice learning refers to learning that is up to the individual (Jacobsen, 2016). It represents the majority of human learning over a lifetime (Falk & Storksdieck, 2005; Rennie & Johnston, 2007). Some estimates state that informal learning makes up as much as 70-90% of a person's learning (Latchem, 2014), or a ratio of 4:1 informal:formal (Cofer, 2000). While free-choice learning can happen at the workplace (Le Clus, 2011), venues such as museums are at the forefront (Falk & Storksdieck, 2005; Rennie & Johnston, 2007).

The first generation of interactive museums started with the Exploratorium in San Francisco and the Ontario Science Centre in Canada. Both opened in 1969 (Patiño, 2013; Schiele, 2008). These centres shifted the paradigm, from communication being a tool in the service of scientific knowledge, to a way to achieve scientific literacy (Schiele, 2008).

All types or learning are valuable and contribute to an individual's cognitive, emotional, and social growth (Eaton, 2010). Mutual support between formal learning and nonformal and informal learning ensures competences needed in rapidly-changing modern societies (Tudor, 2013). Under a framework of lifelong learning, "the validation of nonformal and informal learning outcomes should be promoted" (European Parliment Council, 2008, p. 2). It has been challenging to validate nonformal and informal learning (Broek, Buiskool, Van Oploo, & De Visser, 2012), but by 2010, 22 countries were already involved in the

OECD’s mission to recognize outcomes from nonformal and informal learning (Werquin, 2010).

1.4.2 Informal learning: fun or fact?

The White Oak Institute (2016) is an international museum planning firm. Jacobsen (2016) analysed outcomes that are most sought by museums with which the White Oak Institute has collaborated (most based in USA). Learning was second. Revenue was first. For some, the need to increase visitor numbers has gone too far. Science centres have been criticized for promoting fun and enjoyment rather than education and science (Fors, 2006).

Visitors arrive guided by their own interests, and they learn either what they choose to or what resonates with their previous knowledge or experiences (Falk & Needham, 2013). Learning also depends on a visitor’s values, attitudes and affect (Longnecker, 2016). It is valid if learning science is not in a visitor’s free-time radar (M. Burns & Medvecky, 2016), even if science learning is the venue’s primary objective. Nonetheless, the allegation that one does not learn while having fun is not based on personal interests, but on perceptions. Students tend to reject science subjects (Stodolsky, Salk, & Glaessner, 1991; C. Williams, Stanisstreet, Spall, Boyes, & Dickson, 2003) because they consider them difficult or irrelevant (C. Williams et al., 2003). This gives the false idea that learning science while having fun is counterintuitive.

Even though there is no consensus agreement on precisely what learning science means (Dicks, 2013), there are various views about what learning science entails (Longnecker, Elliot, & Gondwe, 2014). As previously explained, learning science (increasing scientific literacy) involves not only factual knowledge, but engagement and self-beliefs. Moreover, visiting a science centre is highly correlated not only with science knowledge and understanding, but also with interest and curiosity in science, as well as with confidence in and identification with science (Falk, et al., 2016).

Scientific literacy is significantly correlated with enjoyment (Lin, Hong, & Huang, 2012). Some young people don’t like science because they find it boring (Linder et al., 2010; New Zealand Office of the Prime Minister's Science Advisory Committee, 2011). Therefore, interest and motivation are key elements in learning outside of school (Boekaerts, 2010; Sefton-Green, 2012). People can get enthused if science is taught in places that explain science in new and exciting ways (Braund, 2012). “If museums are places that support the public’s learning, the gateway to that learning is interest” (Falk & Dierking, 2016, p. 93).

Enjoyment of science is central to predicting students' participation in science activities (Ainley & Ainley, 2011). Interest in knowledge is also a driving force behind formal scientific research (Krapp & Prenzel, 2011). Enjoyment and interest have complementary effects; when both are present while working on a science topic, individuals engage with that topic and learn science (Ainley & Ainley, 2011).

Knowledge acquisition by the public is complemented by other important benefits, such as the opportunity to encounter science in an exciting or inspiring way (Jensen & Buckley, 2014). Museums are stimulating places to explore the world of science and technology (Ramey-Gassert, 1997). Visitor engagement is first experienced as expectation, because they often attend expecting to satisfy intellectual needs (Falk & Dierking, 2012). Depending on other constrains, they will stay as long as the place continues to keep them engaged (Serrell, 1997).

Experiences from a wide variety of informal environments do more than provide enjoyment and engagement; they provide substance upon which more systematic and coherent conceptual understanding and content structures can be built (National Research Council, 2009). Math and science out-of-school activities influence youth expectations and values; in turn, these influence the high school courses youth choose to take (Simpkins, Davis-Kean, & Eccles, 2006).It takes engagement, motivation and belief in self-capacities to learn and succeed (Christenson et al., 2012; Schunk & Mullen, 2013). The ultimate engagement is known as *flow*, a deep involvement where people spend time doing something seemingly effortlessly (although it actually requires high concentration), with no other reward than satisfaction of doing (Csikszentmihalyi, 1990).

There is an important clarification to make, here. Some authors take it so far as to suggest that enjoyment and learning should be recognized as both precursor and benefit, respectively (e.g., Colliver & Fleer, 2016; Grinell, 1988; Lillard et al., 2013; Rogers, 2013; Wood & Attfield, 2005), or even to consider that for children there is no valid distinction between playing and learning (e.g. Wellington, 1990). However, "fun may be a necessary but insufficient condition of learning in an interactive science gallery" (Archer, Dawson, Seakins, & Wong, 2016, p. 930). In this research, science engagement and scientific knowledge are not considered two sides of the same coin; each is one aspect of scientific literacy (see Figure 1-2). "Learning and fun are not opposite ends of a single continuum but independent dimensions" (Renner, 2011, p. 114).

1.4.3 The role of exhibits in learning

Science centres influence science learning in their visitors through the exhibits. Thus, exhibits become a crucial element to analyze when studying informal learning at a science centre. In order to learn science from an exhibit, it’s necesarry that exhibits grab the attention and keep visitors engaged with a learning activity long enough (Shettel, Butcher, Cotton, Northrup, and Slough, 1968; Bitgood, 2016). In this sense learning at exhibits shouldn’t be studied only from its learning aims (e.g., inspire passion), but also include what makes a visitor approach (e.g. sensorial stimuli), what keeps them engaged long enough (e.g. interactivity), and how to know that science learning is actually happening (e.g. phenomena exposure).

1.5 Assessment of Learning: the gaps

Diversity and popularity of museums have increased dramatically in the last few decades (Shaby, Assaraf, & Tishler, 2016). There are currently over 55,000 museums in 202 countries worldwide (De Gruyter Saur, 2017). Persson (2015) estimates more than 3,000 are science centres, receiving more than 300 million visitors each year. In the US, museums classified as Science & Technology Museums & Planetariums comprise 3% of the national total (Frehill & Pelczar, 2018).

Museums have been historically oriented towards collections and research, but they have become more and more seen as institutions for public learning (Shaby et al., 2016). With so many science centres around the world, there are plenty of studies related to learning assessment. Several show evidence of increasing scientific knowledge at science centres (e.g. A. Anderson, Bequette, Cosbey, Haupt, & Hughes, 2016; Falk & Needham, 2011; Martin, Durksen, Williamson, Kiss, & Ginns, 2016; National Research Council, 2009). There is evidence of science engagement (e.g. Barriault & Pearson, 2010; National Research Council, 2009; Schwan et al., 2014). There are also studies showing self-beliefs in science increasing due to visiting science museums (e.g. Jarvis & Pell, 2005; Martin et al., 2016; Şentürk & Özdemir, 2014).

However, this work is not complete. Science is about finding what can be improved, and doing it. In this section, the main gaps around assessing learning at science centres are presented.

1.5.1 Assessment in New Zealand

Unfortunately, New Zealand lacks studies related to learning assessment in science centres. New Zealand research on museums is mainly limited to the non-science ambit (e.g. MacDonald, 2018).

What New Zealanders are most proud of is general sporting achievements (41%) (Research New Zealand, 2018). However, 77% of New Zealanders consider innovation, science, and technology very important in creating a sense of a national identity, slightly less than landscape and environment (81%), but more than sports (64%) (Ministry for Culture and Heritage's Cultural Statistics Programme, 2009). Ernest Rutherford outstands as a scientist 11% of New Zealanders mention they are proud of (Research New Zealand, 2018).

The differences between the most and least engaged New Zealand students depend more on out-of-school activities than activities within science classrooms (Woods-McConney, Oliver, McConney, Maor, & Schibeci, 2013). This makes science museums of utmost importance in New Zealand. Actually, "given the speed with which science is changing, these resources [museums and science centres in New Zealand] may become a much more integral part of the formal science education process" (Gluckman, 2011, p. 7).

Research on informal education in countries like New Zealand suggests its impact can be significant, but more evidence is required. Studying facilities offering informal education in New Zealand can contribute to understanding how informal learning happens. This study had the opportunity to examine informal learning at the Otago Museum's science centre, described in Chapter 2. Before this research, all the evidence of science learning at the Otago Museum's science centre was anecdotal (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2016). But, "outreach programs should be evidence-based" (Archibald, 2015, p. 1).

1.5.2 Obtrusiveness

Table 1-5 shows some of the many methods available for assessment in informal settings. The three most popular methods to collect data at museums are interviews, observations, and surveys, in that order (Grack Nelson & Cohn, 2015). All these methods can shed light on learning, but some can be more obtrusive than others.

Although observing how visitors interact is an accepted methodology in assessing learning experiences (Barriault & Pearson, 2010), "[t]he visitor may change their behaviour if they know they are being watched" (Grack Nelson & Cohn, 2015, pp. 29-30).

Yalowitz and Bronnenkant (2009) state that data collectors being noticed by visitors happens ‘very rarely’. However, it is erroneous to think that visitors who explicitly approach the researcher to ask a question are the only ones aware of their presence. The researchers themselves acknowledge following visitors and writing down their observations is obvious. “Writing on clipboards is noticeable” (Yalowitz & Bronnenkant, 2009, p. 53).

Table 1-5
Methods for collecting data for visitor studies in museums

Method	**References (examples)**
Observing visitor behaviour	(A. Anderson et al., 2016; McCubbins, 2016; Pattison et al., 2017; Randi Korn & Associates Inc., 2006; Shaby, Assaraf, & Tal, 2017; Yalowitz & Bronnenkant, 2009; Barriault & Pearson, 2010)
Recording visitor behaviour	(A. Anderson et al., 2016; Dicks, 2013; Hauan, DeWitt, & Kolstø, 2017; Pattison et al., 2017; Zimmerman, Reeve, & Bell, 2010)
Interviews	(A. Anderson et al., 2016; Bequette, Svarovsky, & Ellenbogen, 2011; Cardiel & Pattison, 2014; Randi Korn & Associates Inc., 2006; Zimmerman et al., 2010)
Informal chats during the visit	(Dicks, 2013)
Focus groups with visitors	(Dicks, 2013)
Personal meaning maps	(Falk & Storksdieck, 2005)
Asking students to have a structured visit with specific tasks	(Hauan et al., 2017)

It is also naïve to think that visitors asked to behave normally will do so when they know they are observed or recorded. Unsurprisingly, there is evidence suggesting behaviour can be affected by the awareness of being observed or recorded (Mayo, 2004; McCambridge, Witton, & Elbourne, 2014; Sedgwick & Greenwood, 2015).

Finally, even if the observer manages to remain unnoticed, observing visitors is a good method to find what people do, but it is not useful to know what messages they are taking away (McKenna-Cress & Kamien, 2013).Visitor interviews avoid the problem of visitors changing their behaviour during the visit, but it can still result in a feeling of being observed. Interviewers themselves become a factor in a visitor’s cooperation (Jäckle, Lynn, Sinibaldi, & Tipping, 2013). Interviewers also influence the respondent’s tendency to agree with the questions regardless of their content (the “acquiescence effect”): “Acquiescence distorts conclusions made from surveys by artificially increasing levels of support for survey questions and changing relationships among survey items" (Olson & Bilgen, 2011, p. 100).

From the three main methods, a researcher has probably the least effect on visitor's behaviour and responses when conducting surveys. Since surveys are anonymous, visitors may feel they can respond more honestly than during in-person interviews (Grack Nelson & Cohn, 2015). Self-reporting still makes visitors aware of the research, but it is a less obtrusive alternative. It also has the advantage of allowing the researcher to collect more data than by observation or interviews. For example, Longnecker et al. (2014) were able to collect two thousand surveys at almost 60 events, a quantity almost impossible to get by other means.

Most of the above methods also intrude on visitors' time. Visitors go to a science museum in their free time, and it is unknown how many of those who agree to participate, do so because they really have the time and the energy, and how many do so out of courtesy. Either way, their responses are pressured by available time and potentially tiredness, especially at the end of the visit. Surveys can also take a lot of time. For example, Thuneberg and Salmi (2018) assessed knowledge gains from a visit to a museum with an one hour pre-test and 30min post-test. Even though the surveys in this case were conducted with students at school one week before the visit and one to two weeks after the visit, 30 minutes and one hour require a large amount of mental work.

Excessive time and obtrusiveness do not necessarily come from a long single method, but from comprehensive research that tries to obtain all data from the same sample. For instance, Zimmerman et al. (2010) investigated the museum visits from the family perspective. First, they conducted a pre-visit interview of 10 to 20 minutes. Then two researchers followed the family, one video recording their activities and one writing field notes. Children wore microphones during the visit. A five to 10 minutes interview was conducted after the visit. In total, the research team spent between 1.5 and 2.5 hours with each family. The sample size consisted of 44 individuals in 15 families.

1.5.3 Likert-type scales: proved but improvable

Likert-type scales are a popular method to collect self-reports in surveys, but they can present issues that need to be acknowledged. First, several studies remove the middle neutral option (e.g. Cardiel & Pattison, 2014) out of concern for the validity of such responses. Masuda, Sakagami, Kawabata, Kijima, and Hoshino (2017) suggest excluding respondents who select the middle option from statistical analysis, because they might be inattentive or unmotivated, and therefore respond carelessly.

Nadler, Weston, and Voyles (2015) compared the ordinal ‘Neither’ with a nominal ‘No opinion’. ‘Neither’ was more frequently used in a 5-point Likert-type scale (not including ‘No opinion’) than ‘No opinion’ in a 4-point Likert-type scale (not including ‘Neither’). They concluded that the more frequent use of ‘Neither’ shows the option represents more than ‘No opinion’. What Nadler et al. (2015) and others tend to ignore is that the midpoint has two possible meanings: undecided and truly neutral (Raaijmakers, Van Hoof, 't Hart, Verbogt, & Vollebergh, 2000). By removing the midpoint as an attempt to improve reliability, researchers bias the results. They not only prevent respondents from choosing a valid option, but, when removed, participants tend to select options from the positive side of the scale (Worcester & Burns, 1975). Additionally, it is naïve to think that forcing undecided respondents to choose a non-neutral option means they have made up their minds, rather than making a ‘forced, yet undecided’ choice.

In addition to the biases mentioned above, removing the midpoint can affect the ability to use parametric tests on such even-numbered scales (see Appendix E for a discussion).

> The possibility of a double meaning of the midpoint ─ even when a separate ‘don’t know’ category is offered ─ is not an argument for simply withdrawing these response categories. On the contrary, it is advisable for scales to include both, a separate response category ‘don’t know’, and a midpoint. (Raaijmakers et al., 2000, p. 213)

‘Don’t know’ (or ‘No opinion’) is also a legitimate response (Fowler, 2013; Krosnick et al., 2002; Pearce-Morris, Choi, Roth, & Young, 2014; Raaijmakers et al., 2000) that Likert-type scales frequently leave out. This forces undecided respondents to choose a random answer, biasing the survey results and affecting the instrument’s validity.

The use of emojis can also affect the reliability of data collected from Likert-type scales. Smiley faces are often used to help children interpret the scale (Hall, Hume, & Tazzyman, 2016; Read, MacFarlane, & Casey, 2002; Reynolds-Keefer, Johnson, Dickenson, & McFadden, 2009). The typical range of sad/angry to fun/happy faces may be a good option to rate enjoyment (Longnecker et al., 2014), but not in all cases. If sad/angry faces are used to measure ‘disagreement’, the faces may bias responses, as children tend to select the positive ratings depending on how enjoyable they expect the experience to be, not on the particular statement (Hall et al., 2016). A better set of symbols is needed to represent a continuum from strongly disagree to strongly agree.

Even though it is inherent to Likert formats, the use of labels to describe the anchors may pose a threat to considering them as interval scales, instead of merely ordinal. This is because the distance between points is not necessarily perceived as equidistant (Lantz, 2013).

1.5.4 Beyond the self-report: objective testing

Assessment in informal environments has heavily relied on self-reporting (Dunning, Heath, & Suls, 2004; National Research Council, 2009). Often, it is assumed that an honest respondent is enough for an accurate self-report (Paulhus & Vazire, 2007). However, this is not so. For example, highly literate people with the imposter syndrome may feel insecure of their knowledge, capacities and achievements (Parkman, 2016; Sherman, 2013). On the other end of the spectrum, people with low literacy may be unaware of their own ignorance and overestimate themselves in what is known as the Dunning-Kruger effect (Dunning, 2011; Schlösser, Dunning, Johnson, & Kruger, 2013). “People overestimate themselves. They hold overinflated views of their expertise, skill and character” (Dunning et al., 2004, p. 72).

> In most of the cases, low-performers overestimate their skills in self-assessments. These findings have implications for theory and practice. As *self-efficacy scales* [emphasis added] are not reliable instruments to assess IL [Information Literacy] skills, knowledge and skill tests and practical assignments may be used for the important task of IL assessment. (Mahmood, 2016, p. 207)

The ‘familiarity hypothesis’ considers that an individual’s familiarity with a science topic is a good reflection of their actual factual science knowledge[11] (Ladwig et al., 2012). However, despite its popularity, it may not be accurate; not only highly and lowly literate people may be susceptible of misreporting their knowledge. Respondent’s confidence is based on the ease with which potential answers come to mind, making people genuinely believe their knowledge or understanding is correct if they feel familiar with it, irrespective of whether it is right or not (Kelley & Lindsay, 1993; Mbewe et al., 2010; W.-C. Wang, Brashier, Wing, Marsh, & Cabeza, 2016). This bias is called the ‘illusory truth effect’. For instance, Mbewe et al. (2010) showed that the familiarity about science process skills that pre-service teachers self-reported had a mean of 34.6, but their conceptual understanding objectively measured was only of 19.1.

As it can be seen from the above, without demeriting self-reports, scientific knowledge needs to be objectively measured. Unfortunately, using formal testing to measure knowledge is a frowned-upon methodology in informal settings. “Arranging for tests before and after the experience or setting up other traditional measures in many museums and science centres can be disruptive, or even inappropriate” (Fenichel & Schweingruber, 2010, p. 104).

[11] It also considers that this familiarity is positively correlated with science support, but this is beyond the scope of this research.

Unlike students in a classroom, visitors shouldn't worry about grades (Ham, 2016). The author acknowledges that "if not carefully designed, assessments of content knowledge can make learners feel inadequate, and this throws into question the validity of the assessment" (National Research Council, 2009, p. 63), and that using "textbook-like questions… to judge the nearness of an individual's answer to the expert's version of the scientific story … is a limited approach to documenting what people understand about the world around them (National Research Council, 2009, p. 63). However, not assessing content knowledge objectively, and limiting the assessment to self-reports and indirect measures that work under untestable assumptions, is also a limited approach. Very few recognize the importance of objective measures besides self-reports.

Bozdoğan and Yalçın (2009) are amongst the few to use a pre-test post-test methodology with informal learning, with a school group visiting at a science centre. They randomly selected 31 8th graders from a school (also selected randomly). These students filled out an 'Academic success test' before the visit to a science centre, as well as one within a week after the visit, and another four weeks later. One multiple-choice example question from the test was 'Which of the following are the structures that swallow objects in space in an irreversible manner? (Black hole / Supernova / White dwarf / Black dwarf)'. The mean numbers of correct answers were 6.3 on the pre-visit test, 9.4 for the one-week post-test, and 9.8 for the five-week post-test.

Cigrik and Ozkan (2015) also conducted a pre-test, post-test study. They also studied school children (12 to 14 years old, 6th grade from the same school), but included both an experimental and a control group (N=25 each). The experimental group visited Bursa Science and Technology Center every two weeks for six weeks in total. The control group continued with their regular learning at school. Before starting the visits, the experimental group (M=11.7 correct answers) scored no differently from the control group (M=11.2), but by the end of the study, scores of the experimental group (M=19.1) were significantly higher than those of the control group (M=9.2). It is not reported if the decrease in the control group was significant or what could have caused it.

Martin et al. (2016) studied the effect of a self-paced, school vacation programme (situated in a medical science museum) on enhancing content knowledge in 167 elementary and secondary school students aged 10-16 years. Participants completed online tests both before and after the programme[12]. The programme used a zombie theme to teach health

[12] How long before and after they were answered was not reported. Since tests were voluntary and at home, it seems there were no restrictions on when to answer them.

concepts. Content knowledge was assessed with 14 True/False items (e.g., ‘Cells such as T-cells and B-cells are what the body uses to fight off foreign invaders’). Content knowledge increased significantly from M=75.1% to 77.2%[13] correct answers on the pre and post-tests, respectively.

Although their goal was not to measure the change in knowledge, but rather to find predictors of it, Thuneberg and Salmi (2018) compared the results of a (one-week) pre-test and (7-13 day) post-test for 2,591 6th grade students from four countries who visited science centres.

A study conducted by Salmi, Thuneberg, and Vainikainen (2015) assessed knowledge in mathematics before (one week) and after (7-11 days) a visit to an exhibition. The exhibition was designed in Estonia, but data were collected in Latvia (n=408) and Sweden (n=665) from 6th graders (12 to 13 years old). The study was not concerned with assessing the pre-post differences, but finding how post-visit knowledge correlates with pre-visit knowledge, self-concept, and mathematical thinking skills, as well as finding cross-country differences. Thus, although they measured pre and post knowledge, they focused on their goals, not on the pre-post difference.

Finally, no studies have been found where scientific knowledge was measured objectively at a science centre with regular visitors. It is to be clarified that acknowledging the gap in this respect does not mean other methodologies are discredited. On the contrary, the author recognizes the value of them and filling the gap would bring validity to methodologies that are based on postulates. For example, Serrell (1997) suggested that visit time may be a measure of learning and, based on this assumption, created some indexes. This methodology is unobtrusive and simple, and has been used in this research. Measuring scientific knowledge concurrently with visit time, could help validate that methodology.

1.5.5 Pre-beliefs, post-beliefs

Of the three areas of scientific literacy, self-beliefs in science is the least studied. Most of the research about self-beliefs before and after an intervention found in the literature was conducted in formal education settings. There is little research on the effects of informal learning beyond the classroom environment.

Studies related to self-beliefs and assessed before and after visiting an informal setting are mainly limited to school trips and focusing on other constructs different from the specific

[13] Significance is not related to the changes in mean, but to the consistent change in most of the participants. So, a consistent small increase can be significant, and big random changes may not be. See Chapter 3 for more on statistics.

self-beliefs considered in this thesis. For example, attitudes towards nature and biology topics (Ballouard, Provost, Barré, & Bonnet, 2012; Kamarainen et al., 2013; Prokop, Tuncer, & Kvasničák, 2007; Sturm & Bogner, 2010) and attitudes towards science, but different from the self-beliefs considered in this thesis (Holmes, 2011; Jarvis & Pell, 2002).

For example, Martin et al. (2016) studied the change in self-efficacy in 167 elementary and secondary school students (from 10 to 16 years old) participating in a program hosted by a health and medical science museum. Responses to two items: 'If I try hard, I believe I can do well in science subjects that involve human biology' and 'If I don't give up, I believe I can do difficult schoolwork in science subjects that involve human biology', were collected during a period of 20 days of vacation periods. Data came from voluntary participants that arrived for the museum program during school holidays. Each question was in the form of 7-point Likert-type item. The score increased significantly from M=5.7 to 5.9. However, as it will be discussed in Section 1.5.6, one item is related to self-efficacy, while the other one to self-concept.

Şentürk and Özdemir (2014) examined the effect of visiting a science centre on attitudes towards science in students 11-14 years old. The instrument was administered before, immediately after, and one week after the visit; its scale had six constructs, including 'Self-concept in school science'. The study compared a control group (N=46) to an experimental group (N=46). Participants in the experimental group visited the Middle East Technical University's Science Centre (Turkey), while the control group continued with regular activities in their school. Self-concept almost didn't change in the control group, while the experimental group increased from approximately 25.5 points[14] before the visit to over 29.1 immediately after the visit, then decreased to a still significantly higher 27.9 one week later.

Kind, Jones, and Barmby (2007) developed an instrument to measure self-concept in science. Their goal was not to was not to measure pre-post difference. However, to test their instrument, the questionnaire was completed by 932 students (aged 11-14) two weeks before a visit to Lab in a Lorry (a mobile laboratory), and by 668 two weeks after. The questionnaire was not pre-post matched, and not all students who completed the pre-questionnaire visited the laboratory. It is not clear if those pupils who didn't visit Lab in a Lorry did so because they were not interested or how they were selected. Self-concept in science decreased from M=3.41 to 3.24. It is unknown why it decreased and if the difference is significant, since the

[14] This value is not reported in the article. It was estimated from the figure by increasing its size and interpolating the value according to the mark. The closes value to the mark with one decimal point was 25.5.

authors didn't report anything else related to this unexpected decrease. It may be due to the lack of control on the pre and post populations.

The last case of a pre-post assessment of self-belief comes from Sasson (2014). The researcher used a 13 item, five-point Likert-type scale to assess students' attitudes towards science and self-efficacy before and after their participation in a science centre programme. It is not clear which, or how many, items were used to assess self-efficacy. Scores before (M=4.03) and after (M=4.05) the intervention did not change. This may be due to not matching pre and post, as the pre has a sample size of 745 and the post only 475, and this difference is not explained.

1.5.6 Self-concept and self-efficacy: related, but not the same

Self-concept and self-efficacy are closely-related constructs (Bong & Skaalvik, 2003; Jansen et al., 2015). They are commonly associated with the confidence that allows individuals to embrace difficult tasks as challenges to be mastered and, if they fail, to recover more quickly. This makes confidence a critical component of school success (Pajares & Schunk, 2001).

Some authors have used the terms interchangeably (e.g. Wender, 2004). However, they are not the same. Self-concept is a self-belief of an individual as a whole—what they ultimately think they holistically *are* (e.g., How do I feel about myself as a runner?). Self-efficacy is a self-belief of what they *can achieve*, based on their own abilities, in some specific domain (e.g., Can I run 100 metres in under 12 seconds?). The relatively malleable and future-oriented self-efficacy acts as a precursor to the fairly stable and past-oriented self-concept (Bong & Skaalvik, 2003). Note how running 100m under 12s requires an assessment that doesn't depend on peers' capacities, but on one's own. Conversely, assessing oneself as a runner depends on a broad range of factors, including peers' performance. Understanding the differences between these two terms (and other self-beliefs) is crucial to assess them correctly.

For example, it was stated previously that Martin et al. (2016) studied self-efficacy with two items. The item 'If I don't give up, I believe I can do difficult schoolwork in science subjects that involve human biology' measures self-efficacy, as it implies a specific task (doing schoolwork). However, 'If I try hard, I believe I can do well in science subjects that involve human biology' does not measure self-efficacy, but self-concept. How well an individual perceives he/she can do in science is something broad that is completely contextualized. This is probably why the measured change is significant, but small.

Şentürk and Özdemir (2014) do not provide any details of their instrument, but to the original source (Kind et al., 2007) reveals a 5-point Likert-type scale. The construct self-concept in science[15] comprises 7 items: I find science difficult / I am just not good at Science / I get good marks in Science / I learn Science quickly / Science is one of my best subjects / I feel helpless when doing Science / In my Science class, I understand everything. Most of the items are indeed related to self-concept, but ‘I get good marks in Science’ is not even a self-belief. It is a fact; it is on a report card. If that item changed after the centre visit, then a hidden variable is affecting the results.

1.5.7 Whose opinion?

Studies usually focus on opinions from random visitors, looking for what typical visitors take away from their visit. However, there is no such thing as a ‘typical visitor’ and not all kinds of information can come from surveying visitors. One perspective that is often missed is that of staff. Science centres do not magically appear one day and start working by themselves, it’s in decision makers’ hands what exhibits should be included, and it is science communicators (floor staff) who decide how to interact with visitors in order to help them understand the scientific principles. Therefore, how staff perceive science and what are their guiding principles of science communication, become a crucial element in understanding how visitors learn science at science centres.

However, even knowledgeable and committed staff cannot know how visitors are going to interact with the exhibits and what science they are going to take away. For example, it is usually claimed that hands-on interactives are engaging and good for learning, but Holstermann, Grube, and Bögeholz (2010) investigated 28 typical hands-on biology activities and found that only seven had a positive effect on interest; one even had an adverse effect. What were the factors that went into producing an overall positive effect in seven activities, and what created an overall negative experience in one? It may be a difficult question for staff, whose knowledge of science puts them in a different scenario than that of visitors when coming to understand what a ‘typical’ visitor is taking away. Asking the same question to visitors who are just attending the science centre to have a good time may also be too much, not only because of matter of time, but because not all may have tools to understand and dissect the science behind an exhibit.

The solution to these crossroads was young people who were highly-involved with science while still keeping the essence of being a visitor. “[Engaging young people in

[15] Notice that they call this construct ‘self-concept in science’, not ‘in school science’.

designing exhibits] provides the involved students with a unique vantage point from which to observe the institutional relativity of scientific knowledge" (Achiam, 2019, p. 46). Section 3.3.2 expands on how these children performed an in-depth analysis of what visitors might be taking away from the exhibits and how to make it closer to the goals of staff.

Coming back to the surveys, another frequent gap in studies about learning at science centres comes from the almost exclusive focus on children. If children are the future, adults are the present, and adults' lifelong learning can also be influenced by visiting science centres (Gutwill, 2018). Studies that overlook adult learning, or treat it as a collateral result while studying children, are missing valuable information. Only a few researchers have paid full attention to the effect of science centres on adults (e.g. Falk & Needham, 2013; Heimlich & Horr, 2010).

Even when the focus of an adults visit is to engage children with science, parents' attitudes towards science have a large role in cultivating children's own attitudes towards science (Alexander, Entwisle, & Olson, 2007; Gunderson & Levine, 2011; Halim, Rahman, Zamri, & Mohtar, 2018; OECD, 2013b; Perera, 2014). Therefore, engaging parents is a crucial factor for engaging children.

1.5.8 Time as a measure of learning

It has been proposed that visit time is not only a measure of engagement, but of learning (Borun, 1998; Serrell, 1997, 2010). However, although evidence has been shown indicating it, the author found no studies that assess scientific knowledge objectively and compare its gains with visit time.

1.6 Uncovering Learning

1.6.1 Research aim, objectives and research questions

Finding the impact of a science centre is a complex matter that goes beyond collecting demographic data about its visitors (National Research Council, 2009; Rennie & Johnston, 2007). The aim of this thesis is to contribute to the understanding of how learning takes place at science centres. To achieve this, the first objective was to investigate reliable forms of assessing scientific literacy. This inquiry was guided by the first research question (Table 1-6). The second objective was to examine whether changes in scientific literacy happen during a visit to a science centre. If so, what changes occurred (second research question)? The third objective was related to exploring the influence of gender and age in scientific literacy (third research question). The fourth objective consisted of describing some characteristics of science exhibits that could potentially drive visitor learning (fourth research question).

Table 1-6
Research questions

Research question one: Can the scientific literacy of science centre visitors be reliably measured? If so, how?
Research question two: Can scientific literacy be influenced by visiting a science centre? If so, what aspects are affected?
Research question three: Are changes in scientific literacy influenced by age and gender? If so, how?
Research question four: What characteristics of science exhibits influence visitor learning?

Exhibits at Otago Museum's science centre present a great range of topics, from botany to psychology. To obtain a more focused understanding in the objectives above, data were mainly collected from exhibits related to light and electromagnetism, i.e., belonging to the Light Zone or Unseen Forces Zone. The topic of Light, which is the main topic in both zones, is rich in physics phenomena, and provides an excellent opportunity to study both misconceptions and conditions for learning to occur (Feher, 1990; Feher & Rice, 1988). Exhibit names are italicised for sake of identification.

It is important to note that this investigation goes beyond evaluation to include assessment. Evaluation and research are commonly confused because many times they use the same methods, but the goals are different (Patiño, 2013). Evaluation involves approaches and techniques used to make judgments about a given instructional program, approach, or treatment, improve its effectiveness, and inform decisions about its development. Assessment is the set of approaches and techniques used to determine what individuals learn from a given instructional program. Assessment targets what learners have or have not learned, whereas evaluation targets the quality of the intervention (National Research Council, 2009).

Assessment and evaluation approaches have been used in this research to investigate phenomena of science learning with the purpose of exploring how and when science learning takes place in a science centre.

1.6.2 Thesis outline

The thesis is presented in eight chapters plus appendices.

Chapter 1. Informal Learning of Science. Introduction to informal learning and research aims.

Chapter 2. Otago Museum Science Centre. Description of the case study and aims for redevelopment.

Chapter 3. Methodology. Describes the research paradigm, data collection methods used and analysis conducted.

Chapter 4. Scientific Knowledge. Results and discussion about the effect of visiting a science centre on visitor scientific knowledge.

Chapter 5. Self-beliefs in Science. Results and discussion about the effect of visiting a science centre on visitor self-beliefs in science.

Chapter 6. Science Engagement. Results and discussion about the effect of visiting a science centre on visitor science engagement.

Chapter 7. Exhibits and Science Learning. Exhibit characteristics that can influence visitor learning.

Chapter 8. Concluding Remarks. Concluding remarks.

1.7 Summary

This chapter introduced the foundations upon which the rest of the thesis is based. The global challenges of the modern world require scientific knowledge and action. But, this action can only come through the motivation of scientifically informed and engaged citizens. Science centres are a key mechanism for citizens to become aware about science issues and to learn science. Therefore, understanding how learning happens at these venues is crucial. One such place is the Otago Museum’s science centre, described in the next chapter.

To help solve the overarching question of how learning happens at science centres, an integrated framework was developed and presented. In this framework, learning is considered as an increase of scientific literacy, and literacy is in turn comprised of scientific knowledge, science engagement and self-beliefs in science. How this learning can be assessed was presented.

The gaps in previously used methodologies were discussed. From them, it can be concluded that data collection from visitors should, ideally, be unobtrusive and short. Likert-type scales require several characteristics that usually are dismissed or removed, which creates a resulting bias. Some of the ideal characteristics need to be integrated, such as creating a new set of more suitable emoji and, ideally, producing a scale with no verbal anchors. Few studies found in the literature have conducted pre-post formal tests at science centres. Even with school children as participants, the number of studies is minimal. It is possible that the strong rejection of ‘school-like’ tests is only based on theoretical fears, rather than on actual unsatisfactory results. Since scientific knowledge is a cornerstone of scientific literacy, it makes sense to objectively measure how actual knowledge is updated during a science centre visit. self-beliefs in science are still rarely studied. Only two studies were found that tried to compare how self-beliefs change due to visiting a science centre, and both have issues with separating self-concept from self-efficacy. It is necessary to build on what is known about these two constructs to assess them before and after visiting a science centre, and to do so with proper scales.

Complementing these methods, research with engaged participants can contribute important insights about learning in science centres. The next chapter describes the case study used in this research, the Otago Museum science centre. Chapter 3 describes the data collection methods that best suit the research in this thesis and their associated analysis methods.

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

2.1 Introduction
2.2 Museum Staff
2.3 Redevelopment: from Discovery World to Tūhura
2.4 Māori Conversation
2.5 Summary

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 2: OTAGO MUSEUM'S SCIENCE CENTRE

2.1 Introduction

The previous chapter established the basis of this research by describing what scientific literacy is, why it is important, and what aspects might be expected to be modified by visiting a science centre. To unveil if and how science centre visits effectively modify scientific literacy in visitors, the Otago Museum's science centre was used as a case study. The centre was completely redeveloped during the research, changing name from Discovery World (pre-redevelopment science centre) to Tūhura (redeveloped science centre). Data come from both Discovery World and Tūhura, strengthening the conclusions based on mutually reinforcing results, and fostering a deeper discussion when results from both science centres challenged each other.

Staff are vital in a science centre. "Trained interpreters can help visitors make sense of exhibit topics and by that have a transformative effect on visitors' learning" (Schwan et al., 2014, p. 80). If we aim to truly understand how science communication happens at a science centre, we need to know the thoughts of the people behind the centre, their duties, and what they understand about science and science communication. Staff thoughts about science, science communication and the redevelopment, were investigated through interviews with floor staff and decision makers before and after the redevelopment. This chapter presents such a preamble so that the results chapters can be better understood.

After hearing from the staff, an overview of the Otago Museum's Discovery World and Tūhura is provided. Given the variety of science topics at both science centres, light and electromagnetism were chosen to focus the study of how scientific literacy changes. The advantages as phenomena-rich topics, that were present at both science centres are discussed. The Dodd-Walls Centre is also introduced as a partner sponsoring light and electromagnetism-related exhibits.

Museums, including science museums, are cultural institutions (N. Simon, 2010), but unlike Discovery World, and most science centres around the world, Tūhura is bicultural. It not only introduces visitors to Western science, but to indigenous knowledge as well. The last section of this chapter is dedicated to this unique feature. Although this thesis will not study Tūhura's bicultural aspect, having the background reasons behind the new design is important when analysing results.

2.2 Museum Staff

2.2.1 Interviews

Interviews provide contextualized and individualized data (Frey, 2018). Semi-structured interviews (see Appendix A) were conducted with 14 staff members. The interviewed staff can be divided into two groups. Decision Makers (DM) who make high-level decisions and Science Communicators (SC) who have direct contact with visitors. Two supporting staff were also interviewed: the Otago Museum's Māori curator, and a technician who fixes Tūhura's exhibits. To avoid identification, the Māori curator was included in the DM group (n=6) and the technician in SC (n=8). Seven of the interviewed staff members were male and seven were female.

Four science communicators and five decision makers were interviewed twice, before and after the redevelopment. One SC was only interviewed before the redevelopment. Two SC were only interviewed after the redevelopment, but they also answered corresponding questions of the pre-redevelopment interview. The Māori curator and technician only answered questions relating to their areas of expertise, after the redevelopment.

To protect respondents' anonymity, quotes include an identification code, with a number indicates the number indicating the individual participant in the given group. For instance, DM2 means Decision Maker 2; SC5 means Science Communicator number 5. More on how interviews were analyzed in section 3.3.1.

2.2.2 Scientific worldviews

A science centre comes to life not by the devices it contains, but rather the energy of people both in front and behind the scenes. Knowing about the scientific worldviews of those involved, and their expectations for a redevelopment, helps to form a better idea of how the science centre came to be what it is. "[I]n order to fully understand the visitor's experience in the museum, its pedagogical staff's input about the exhibits must be taken into consideration" (Shaby et al., 2016, p. 360).

It is not usually up to the public to make decisions related to museums; it is the Decision Makers who usually filter the public's thoughts and desires (Bandelli & Konijn, 2013). For this reason, their own vision is highly influential. For instance, Decision Makers were in charge of choosing the science exhibits currently on display at Tūhura. But a selection process is not arbitrary—it depends on what they consider science to be, what they feel should be conveyed, and how.

Once in place, Science Communicators can become an exhibit's living voice. "An explainer is the human interface between the museum and the public" (Rodari & Xanthoudaki, 2005, p. 2). Facilitators serve an important supporting role, not only because they can enhance learning (Fenichel & Schweingruber, 2010; Rodari & Xanthoudaki, 2005), but because their communication skills may allow a visitor to leave enthused (Rodari & Xanthoudaki, 2005). What and how a Science Communicator chooses to explain also depends on personal scientific worldviews. In certain cases, visitors will defer to the information they share. For example, Kamolpattana et al. (2015) explain that in Thai culture, people in lower social position are encouraged not to disagree with those in a higher position. This is known as the *Krang Jai*— a situation that may appear in some circumstances.

Otago Museum staff gave their opinions on what science, science communication, and science outreach are. They also commented on their duties and what they expected of the redevelopment. Even though Science Communicators and Decision Makers may sometimes differ on their opinions, they generally had points of convergence around science topics. This is not surprising, as staff attitudes and behaviour may be influenced by the experiences at settings like science centres (Groff, Lockhart, Ogden, & Dierking, 2005).

Most of the interviewees were highly educated; almost half hold a PhD, and having studied science is a current staff requirement. Those who don't hold a degree in science have a tangential involvement with the science centre, or entered the team before it became a requirement. "I'm very lucky to be doing what I love, not what I studied [chuckles]" (SC3).

Decision Makers see themselves primarily as decision makers, and some of them also as scientists. Science Communicators mostly described their duties in terms of running the science centre and interacting with visitors, followed by being transmitters of enthusiasm and facilitators of learning. "As a Science Communicator, my job is to make people excited and to teach people about science" (SC4).

Staff consistently described *science* as a tool to understand the world. "Science is the best tool that humans have developed to explain the universe around them" (DM3). "It's a way of trying to make sense of the world in an objective manner beyond our very poor senses" (SC6). It is driven by curiosity. "I think science can fail only when they stop asking questions" (SC4). To staff, science is methodological. "Science is a method of inquiry where you wanna know about something. And so, you take methodological steps to try to find what it is" (SC3). Science is objective and testable. "[It's] filling knowledge gaps in an objective and verifiable fashion" (DM2).

Opinions diverged slightly when discussing the reach of science. For most, science was seen as universal. “It doesn’t matter what religion you might follow. It doesn’t matter where you are in the world” (SC5). But for some, it is only one of a number of ways of explaining the world. “It’s a belief system. Some people see the world through the eyes of their religion, whatever your religious belief is. Scientists see the world through science. So, I think, it’s the way that you filter and perceive the world” (DM5). Acknowledging the many ways knowledge is acquired is important. Learning science is a cultural activity; it is connected to language and discourse styles (National Research Council, 2009). “When people enter into the practices of science, they do not shed their cultural world views at the door” (National Research Council, 2009, p. 217).

Nonetheless, there was agreement that science is the best approach to know the true nature of things.

> Science is the systematic pursuit of reliable knowledge and knowledge being justified true belief. The hard part is the true, and science gives us, at least a method of forming things that we can approximate to truth. Or we can at least test their validity or test their truthfulness, if you like, by being falsifiable (DM4).

Science communication was perceived by staff mainly as a way to make science understandable to anyone regardless of their background. “It’s primarily introducing people to scientific concepts and principles in a way that is approachable to many different ways of thinking” (SC2). Science communication is “the conveyance of sometimes complex ideas in a manner that is suitable for the audience that you are trying to communicate to. So, the distillation of an idea in a manner accessible to your target audience” (DM4).

Staff see *science outreach* “as a branch of science communication, in where an institution, museum, research group, takes science out to the community, rather than running something in their space” (SC6). From the museum point of view, it is “engaging with people who cannot get to the facility” (DM1).

The definitions of science and science communication applicable to this thesis can be found in Table 1-2.

2.2.3 Staff expectations of the redevelopment

Staff had some institutional expectations such as financial sustainability, but learning stood out as a pivotal goal both before and for the redevelopment “The main vision of Discovery World is to create a space that inspires lifelong learning … To make a space that evokes conversation around science limitlessly in every way” (DM1). “[The] goal is really to create a centre that gives the young people of this region a chance to be inspired by science”

(DM3). Jacobsen (2016) found that usually the main goal is revenue, followed by learning. Here, that order is inverted in responses of Otago Museum participants.

A concurrent concern of staff about Discovery World was that the science centre seemed to be aimed at entertaining little children rather than at showing science. “Some of the stuff is quite gawdy, it’s quite childish, and it’s not child-like, it’s gone past that point” (SC4). “I called it like a glorified playground” (SC1). Some specific exhibits were giving that impression. “I feel it’s [*Air Hockey Table*] a really negative impression to have a game like that in a science centre” (DM3). But there was a more general feeling. “There’s lot of things I didn’t like about Discovery World, colour schemes, the noise” (SC7). According to staff, one of the reasons why Discovery World lost its shine was age.

> To give you an idea, I’m 23 and I grew up here in Dunedin, and I remember going there when I was maybe four or five years old. The other day, we were looking at the photos, and things have changed, of course, over like the last decade or so, but not a lot. (SC1)

Exhibits may be new to young children, but adults grew up with them and it gets tiring. “It’s like going to the movies. You get tired of it, you wouldn’t go and see the same movie again” (DM2). Moreover, natural selection didn’t act in determining which exhibits survived for their communicative qualities; instead, sturdiness determined survival. Any exhibit at Discovery World “wasn’t there by design, it was there because everything else was broken” (DM5).

Staff expected Tūhura to bring fresh interactives where the whole family, irrespective of the age of the members, could engage with them and learn science. “We want the centre to also inspire interest in science across the generations” (DM3).

A sign of the need for a redevelopment was that Discovery World became a walkthrough to the Tropical Forest (butterfly enclosure) for many visitors. “It used to be all about the butterflies” (SC6). The redevelopment was expected to change that. “So that you’re not coming to see Tropical Forest or Discovery World, you’re coming to the new science centre and that’s part of the entire offering” (DM5).On a personal level, Discovery World was no longer appreciated by Science Communicators. “To be honest, I don’t really have anything I really love in the current iteration of Discovery World” (SC1). They expected the redevelopment to boost their enthusiasm. “I think the new Discovery World will, personally for me, reinvigorate my sort of enthusiasm for it” (SC1). Indeed, they felt their joy was revived by Tūhura.

> Overall, I’m really proud of the space and that’s something that I had really hoped to come out of the process. Because I feel like with Discovery World I spent a lot of my time either repairing it or apologizing for it or feeling frustrated by it. While there are

occasions that that's true in Tūhura, I would go away and tell my friends and family '[You] should go and check it out. It's actually really cool'. For me, to feel like I can say that truthfully, is something (SC3).

2.3 Redevelopment: from Discovery World to Tūhura

2.3.1 The Otago Museum

The Otago Museum is an institution where nature, science and culture meet. It is located in the city of Dunedin and it is named after the Otago Region in the South Island of New Zealand. In 2018, it celebrated its 150^{th} anniversary (Otago Museum, 2018b).The importance of this museum to the community is reflected in its having more than 350,000 annual visitors[16] (Otago Museum, 2016, 2017, 2018b). This is almost three times the population of Dunedin, 130,000 in 2013 (Statistics New Zealand, 2019). Most of the museum is free, but the planetarium, science centre and special exhibitions are ticketed areas.

2.3.2 Discovery World

In 1991, the Otago Museum opened a hands-on and play-orientated science centre named Discovery World. To keep it fresh, it underwent several updates during its existence. In 2007, a Tropical Forest (a.k.a. Butterfly House) was added and included with the same admission ticket (Otago Museum, 2018a). Located at a latitude of 46° South, Dunedin is cold most of the year. Thus, it is especially understandable that a warm venue full of butterflies was successful. According to staff, the Tropical Forest became the main attraction, especially for adults, while Discovery World's exhibits zone remained attractive only to younger children. A second major science addition to the museum was the Perpetual Guardian Planetarium in 2015 (Otago Museum, 2018a). This time, its ticket and entrance were independent from Discovery World.The last additions to Discovery World included the following: an in-house developed exhibit ─'Torsion Wave'; a free showcase of science demonstrations ─'The Best Science Show in the History of the World Ever!' (Otago Museum, 2015); and new children's activities and an update of painted graphics (Otago Museum, 2017).

Probably motivated by the upcoming closure, Discovery World reached 73,000 attendees in its final year. This was 3,500 more than the previous year (Otago Museum, 2017).

[16] Reports from the Otago Museum run in financial years from 1 July to 30 June.

2.3.3 Tūhura

Discovery World closed in July 2017 to allow a full renovation (Otago Museum, 2017). After four years of planning, five months of construction, and an investment of NZD$2.5 million, Tūhura Otago Community Trust Science Centre (Tūhura, which includes the Tropical Forest) opened on 15 December 2017 (Otago Museum, 2018b). In its first nine months of operation, Tūhura received almost 64,000 visitors (Otago Museum, 2018b). It was expected that, for the 1-year period of 2018-2019, there would be more than 75,000 paid admissions to Tūhura (Otago Museum, 2018a).

These were the main changes to the science centre after its redevelopment:

- **Larger space**. Adjacent to Discovery World, there was a large independent room called Search Centre (containing some free-to-consult computers, bones and books) and a Computer Suite (storage room) that, combined, covered about 159 sq.m. Inside Discovery World a Tropical Forest Discovery Zone (102 sq.m) served as transition between the exhibits and the Tropical Forest. This space displayed the science of the Tropical Forest. Tūhura used these spaces to increase the science centre exhibition area. Not including the Tropical Forest (215 sq.m), planetarium, Beautiful Science gallery (277 sq.m between both), and their common Welcome area (39 sq.m), the exhibits zone increased from 393 sq.m (Discovery World) to 654 sq.m (Tūhura)[17].
- **Integrated Planetarium and Beautiful Science gallery**. Although the planetarium still requires a separate ticket, it shares entrance[18] with Tūhura. Between the planetarium and Tūhura there is now a free Beautiful Science gallery. The layout makes Tūhura look bigger and more integrated with the gallery and the planetarium. Figure 2-1 shows the entrance to Discovery World and Figure 2-2 the entrance to Tūhura. Notice how Discovery World appeared a separate and smaller space highlighting the Tropical Forest, while Tūhura looks bigger. Tūhura, Beautiful Science and the planetarium share entrance and front desk.
- **Science Show**. The Science Show is no longer run in Tūhura, but there is something similar—QuEST for Science. It takes place in other galleries around the museum and it is free to attend.
- **Improved Tropical Forest**. The butterfly house also underwent a redevelopment, but minor. Figure 2-3 shows some images taken at Tūhura's Tropical Forest.

[17] Personal communication with Margot Deveraux, Otago Museum Project Manager. 7 October, 2019.
[18] The first gate is to enter Tūhura / Beautiful Science / Planetarium. The second door is to go inside the Planetarium.

- **New exhibits**. Discovery World had approximately 35 exhibits[19], Tūhura has 46. The few exhibits that were kept from Discovery World were *Plasma Plates* and *Plasma Tubes* (renamed as *Plasma: The fourth state / Wē Hiko* and *States of Matter*, respectively), *Funhouse Mirrors* (now before Tropical Forest, no longer at the exhibits zone), and an unnamed play table with translucent coloured shapes.
- **Mātauranga theme**. Tūhura addresses something most science centres don't even consider: indigenous knowledge. Mātauranga (Māori knowledge) is a key ingredient within Tūhura. Tūhura not only acknowledges the existence of Mātauranga, but considers that it can complement Western knowledge and ways to understand the world.
- **Layout**. Discovery World was colourful and bright, designed for children. Tūhura is more sober and darker, closer to an artistic experience. "Lighting can be poetry that brings the entire exhibition together and sets the experience" (McKenna-Cress & Kamien, 2013, p. 153)
- **Labels**. Discovery World panels followed a format with sections: Name of the exhibit, What to do, What to look for, and What's happening. In Tūhura, instructions and explanations are minimal to allow a freer experience and interpretation.

Having the opportunity to study the Otago Museum's science centre before and after a major redevelopment is a unique opportunity. It is not the goal of this research to directly compare Discovery World and Tūhura. However, having a similar size, and the same venue and staff, this project obtained data that sheds some light on the differences between them.

[19] The Science Show was not counted as an exhibit, as it consisted of information panels with no associated exhibit. *Plasma Room* is counted as four exhibits in this case.

Figure 2-1. Entrance to Discovery World.

Figure 2-2. Entrance to Tūhura (left), Beatiful Science gallery (right) and planetarium (right).

Figure 2-3. The Tropical Forest (Butterfly House).

2.3.4 The Dodd-Walls Centre: from the Light Zone to Unseen Forces Zone

The Dodd-Walls Centre for Photonic and Quantum Technologies (Dodd-Walls Centre, DWC) is a New Zealand National Centre of Research Excellence; it is a collaboration between five universities, and administratively hosted by the University of Otago in Dunedin. Their research focuses on atomic and quantum optical physics and sensing applications (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2018).

The Dodd-Walls Centre has partnered with several institutions to do outreach. For example, it partnered with the University of Otago's Science Wānanga program to bring experiments to marae,[20] with the Otago Central Rail Trail Trust to develop and install the Interplanetary Cycle Trail, and with the Museum of Transport and Technology (MOTAT) to carry out the Street Science Fair (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2018). With the Otago Museum, they participated in the Queenstown's Luminescence festival through Lab in a Box (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2017), and have travelled as far as the Chatham Islands to do outreach (Otago Museum, 2018b).

[20] Places where Māori communities meet throughout New Zealand.

The DWC celebrated the 2015's UNESCO International Year of Light (100th anniversary of Einstein's Theory of General Relativity) by sponsoring an extensive programme of activities, e.g., Illuminating New Zealand, Te Kōanga, Beambox, and a Light and Art Exhibition (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2016). The DWC sponsored the creation of an enclosed space dedicated to light and electromagnetism in Discovery World: the Light Zone (The Dodd-Walls Centre for Photonic and Quantum Technologies, 2016). The DWC also helped sponsor the development of Tūhura (Otago Museum, 2017). Light and electromagnetism exhibits at Tūhura are now part of the Unseen Forces Zone, a large and open space that also includes forces.

2.4 Māori Conversation

The Māori people are the indigenous Polynesian people of New Zealand. "Mātauranga Māori is the knowledge of Māori. It's the perspective and understanding that Māori have that is particular to their culture" (DM6). Tūhura was the first New Zealand science centre to include both Mātauranga and Western knowledge in its conversation about science. The acknowledgement of the importance of Mātauranga Māori in the development of knowledge can be seen from the post-redevelopment name of the science centre: Tūhura, which means 'Exploration' in te reo Māori.

In one study of school children in multicultural classes, science was seen as new and progressive, but culture was perceived as old and basic (Gondwe & Longnecker, 2015). "The value of native knowledge and their beliefs about the natural world have often gone unrecognized; in fact, many people perceive a conflict between native understanding of the natural world and scientific understanding" (Fenichel & Schweingruber, 2010, p. 132). This perception makes the creation of a bicultural science centre a difficult task. "The main challenges were trying to fit some very hard science. Or to see the similarities between very hard science and its parallels within the Māori world" (DM6). However, knowledge and science learning are inherently cultural (National Research Council, 2009), so is science communication (Davies, Halpern, Horst, Kirby, & Lewenstein, 2019). "What those who disregard it [Mātauranga] fail to comprehend is that pūrākau and maramataka[21] is knowledge generated using the scientific method, explained according to a Māori world view" (Hikuroa,

[21] Pūrākau and maramataka are forms of Mātauranga. They comprise codified knowledge and include techniques to investigate natural phenomena empirically (Hikuroa, 2017).

2017, p. 6). As expected, the museum staff found convergence points to start the conversation.

> When you think about some of the tectonic process, like earthquakes and things like that, you know the Western science will know that that's plates moving and all that sort of things. From a Māori perspective, they view earthquakes, and that other sort of movement, as the movement of a particular god. But he's in the ground, so they understood that the movement came from deep in the ground, even though it was viewed as an unborn child, rather than plates moving. So, that's a similarity (DM6).

Tūhura explains the natural world from two different knowledge bases: science and Mātauranga. "It's not just about having Māori language on the side. It's trying to incorporate some of the Māori ideas about the cosmos or about science as a thread running through the exhibition" (DM3). An exhibition space has perceptual consequences on visitor's emotions (McKenna-Cress & Kamien, 2013) and Tūhura was designed to intertwine Māori culture and Western science using three layers (Figure 2-4). The first layer is "a Whakapapa[22] ribbon, which is our ordering principle. It's the story of the world, but it also creates design" (DM5). The second layer is the Māori imagery on the columns. For someone "who is Māori, who connects with the Māori world, might look at that and see a whole lot of other things that we wouldn't see" (DM5). But even for someone without a Māori background, this layer offers some beautiful imagery, and design and decoration elements of an exhibition awakens pleasant and enjoyable feelings and emotions in visitors (Kottasz, 2006). The third layer is the use of Te Reo along with English. It is not just about language translation, but embracing bicultural meaning as well.

Māori and Pacific peoples represent 24% of New Zealand's population (Statistics New Zealand, 2018). New Zealand's science centres are open to anyone, but "simply exposing individual[s] to the same learning environments may not result in equity, because the environments themselves are designed using the lens of the dominant culture" (Fenichel & Schweingruber, 2010, p. 120). The relative low attendance of Māori and Pacific visitors agrees with Dawson (2012), who found that most participants in Public Engagement with Science in Europe seem to be middle class white Europeans, while minorities and low socioeconomic status were largely absent ."Normally, when you go into a science centre, there's nothing there for Māori kids that identifies as Māori, 'This is me, I can relate to this'" (DM6). Tūhura breaks that paradigm. For instance, the company in charge of *Magnetic Sculptures* originally delivered it with iron filings, but staff demanded actual magnetic black sand. "That's got its own kind of cultural significance, because it comes from beaches that are usually on Māori land" (DM5).

[22] Māori genealogy.

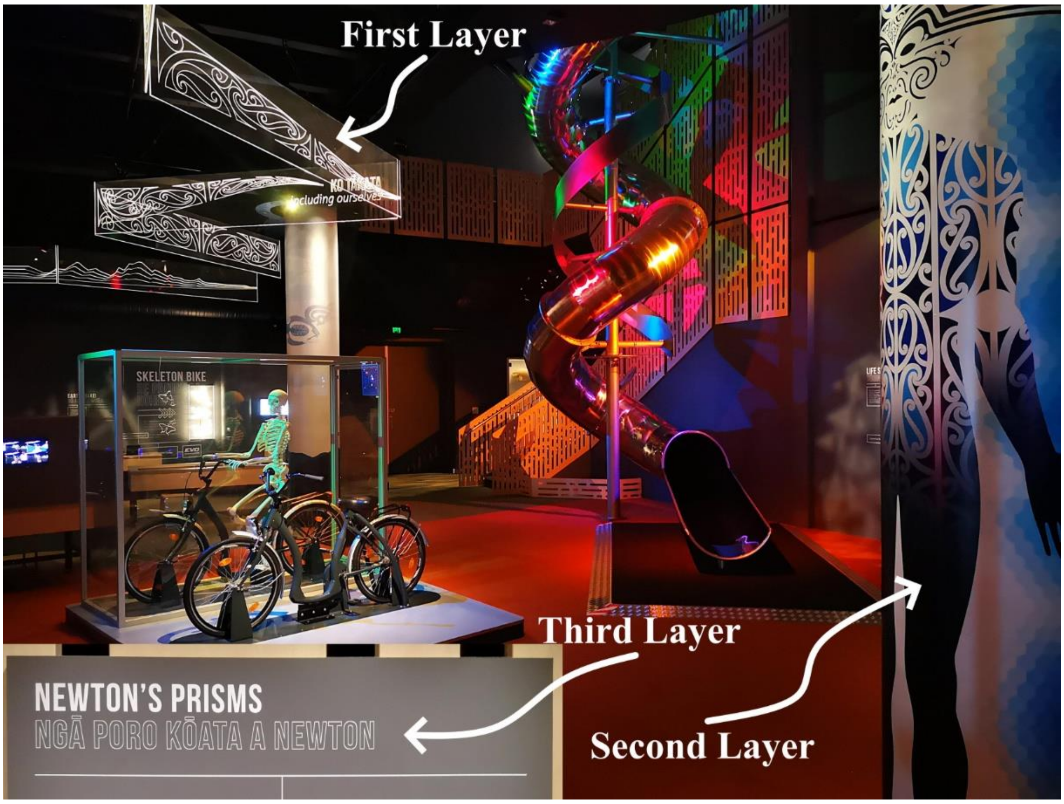


Figure 2-4. Three layers where Māori culture is present at Tūhura. A section of a panel has been superimposed to show the third layer.

Introducing culture is an ongoing process that presents challenges (Longnecker & Scott, 2018) and requires continuous improvement (Fenichel & Schweingruber, 2010). Tūhura is no exception, but its efforts have already been acknowledged by the Museum Aotearoa Service IQ Awards, who awarded Tūhura with the Exhibition Excellence-Taonga Māori award (Otago Museum, 2018b). Visitors embrace this new approach as well, even the younger ones, as demonstrated by the following quote from a 10 year-old visitor. "[I learnt] science, how things work. Learning more about [how things] are cultural and fun for kids". Tūhura provides a model for other bicultural museums. "I would not hesitate to encourage other institutions and places to look at doing the same thing" (DM6). "I'd just wish more science centres in the world did what we are doing here" (SC5).

2.5 Summary

After more than two decades of service, the original Otago Museum's science centre, Discovery World, closed its doors to give way to Tūhura, the current science centre. This renovation allowed the author to compare results from two approaches of science communication (Discovery World and Tūhura).This chapter described both science centres, the history of Discovery World and the new characteristics of Tūhura. One of the main features is the bicultural approach. Mātauranga (New Zealand’s indigenous knowledge) and Western science work together at Tūhura to help visitors understand the natural world from more than one perspective.

The chapter also introduced the Dodd-Walls Centre and its partnership with the Museum, designing and producing exhibits related to light and electromagnetism for Tūhura. A series of interviews with museum staff helped to describe their perspectives about science, science communication and the redevelopment. Understanding their worldviews about science established a base from which we can better understand the two science centres and their effects on visitors’ scientific literacy. Staff see science as a tool to understand the world, and science communication as a way to make science accessible to everybody.Chapter 3 will introduce the research design, methods used to collect data and how these data were analysed.

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

3.1 Introduction
3.2 Research Design
3.3 Data Collection
3.4 Quantitative Data Analysis
3.5 Qualitative Data Analysis
3.6 Summary

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 3: METHODOLOGY

3.1 Introduction

Learning is arguably a science centre's most important goal, but at the same time, it is a concept that refuses to be unanimously defined. Chapter 1 presented a series of research questions that defined learning as a change in scientific literacy, three aspects of which are scientific knowledge, science engagement, and self-beliefs in science. Chapter 2 introduced the Otago Museum's science centre as a case study with two approaches: Discovery World (pre-redevelopment science centre) and Tūhura (redeveloped science centre).

This chapter describes the mixed methods used to collect data to test the research questions. It explains how a combination of quantitative and qualitative methods can assess changes in the proposed three streams of scientific literacy, as well as find the exhibit characteristics that promote learning. All methods were conducted simultaneously with different populations and pre-processed, then analysed with a combination of coding, descriptive statistics, and inferential statistics. How participants were selected is explained, and their demographics are described.

3.2 Research Design

3.2.1 Mixed methods

Research methods can be considered, broadly, as either qualitative or quantitative. Qualitative approaches look for a holistic interpretation of the individual meaning, focusing on depth and richness in context; quantitative approaches look to generalize results by determining relationships among variables (Creswell, 2009; Hernández, Fernández, & Baptista, 2014). But rather than being rivals, qualitative and quantitative studies are ends of the same continuum, and using both provides complementary results which compensate for inherent weaknesses in each individual method (Creswell & Clark, 2017; Hernández et al., 2014).

Given the short amount of time visitors spend at exhibitions, deep learning is not expected (Falk & Dierking, 2016; National Research Council, 2009; Serrell, 1997). However, the possible outcomes of visiting a science centre vary, and can be complex. Combining qualitative and quantitative methods allows researchers to tackle this complexity (Diamond, Horn, & Uttal, 2016; Frey, 2018; Jacobsen, 2016; Procheş, 2016; Sale, Lohfeld, & Brazil, 2002). Using multiple methods is particularly useful when methods include self-reporting (Bjork & Linn, 2006).

Quantitative and qualitative data were collected simultaneously (concurrent mixed methods) in most cases, to allow cross-validation by triangulation (comparison of methods to detect convergence) (Creswell, 2009; Frechtling Westat et al., 2010; Hernández et al., 2014).

3.2.2 One group pre-test-post-test design

Respondents were surveyed before (pre-test) and after (post-test) visiting the science centre to assess changes due to their visit (Creswell, 2009; A. J. Friedman, 2008; Hernández et al., 2014). It is acknowledged that pre-testing may pose a risk of sensitizing and 'cueing' the user, affecting the outcomes (A. J. Friedman, 2008). In other words, visitors may actively search for answers to the pre-test, causing their scientific literacy to increase more than they would in a normal situation. However, matching pre and post responses is a widely-used experimental design that allows for changes to be detected in the same population (Creswell, 2009; A. J. Friedman, 2008; Hernández et al., 2014).

3.3 Data Collection

3.3.1 Interviews

As described in section 2.2.1, semi-structured interviews were conducted with 14 staff members (five decision makers, seven science communicators, one technician and one Māori curator, see section 3.3.4). The protocol and the questionnaires for the interviews can be found in Appendix A.2.

Some questions of the interviews were aimed at solving research question four (Table 1-6), but the rest of the interviews were not addressing specific research questions. Instead, they were aimed at exploring the reationale behind the decisions of what science to communicate and how to do it. This rationale is important to understand, especially when a redevelopment took place.

3.3.2 SWOT focus groups

This method was aimed specifically at solving research question four (Table 1-6) about the exhibit characteristics that influence visitor learning. Notice that staff interviews and visitor surveys complemented SWOT focus groups in answering research question four.

A focus group is a remarkably flexible tool in qualitative research. It allows a group of individuals to discuss, in depth, their perceptions, experiences, and feelings about a topic in a relaxed and informal setting (Barbour, 2008; Fowler, 2013; Frey, 2018; Hennink, 2013; Hernández et al., 2014; McKenna-Cress & Kamien, 2013; Metcalf, Alford, & Shore, 2013; Rio-Roberts, 2011; L. E. Sullivan, 2009).

SWOT (Strengths, Weaknesses, Opportunities and Threats) analyses are typically used in business models (Gürel and Tat, 2017; Leigh, 2009), but they are suitable to be conducted in other types of organizations, including non-profit organizations (Leigh, 2009). Moreover, museums share characteristics of business (see Boylan, 2004; Nobel, 2011) and Jacobsen (2016) even mentions SWOT Analysis as a technique for self-assessment in museum settings (Jacobsen, 2016).

An organization exists in two environments, one internal (itself) and one external. SWOT analysis is a proved technique to make macro evaluations based on positive and negative aspects of the internal and external environments. Strengths and weaknesses are related to internal factors, while opportunities and threats to external factors (Gürel and Tat, 2017).

Given that focus group data can add meaning to SWOT analysis by providing perspectives from other people (Leigh, 2009), the author decided to use them both simultaneously. The thoroughness of the method depends on the time devoted to the task, the number of experts involved and the level of expert consensus (Leigh, 2009). In this new methodology, focus groups were asked to attend the science centre for 30-40 minutes—their time focused at the Light Zone exhibits (Discovery World) or the Unseen Forces exhibits (Tūhura). Each participant carried a sheet of paper with four sections: Strengths, Weaknesses, Opportunities, and Threats (SWOT). Participants wrote down SWOTs for these exhibits and the regions as a whole. Afterwards, participants and facilitators went to a quiet room to have the SWOT focus groups. Using their individual SWOT notes as reference, participants created a group SWOT Analysis. The group discussion was audio-recorded for later transcription.

All participants were allowed and encouraged to express their opinions, whether they agreed with other participant's perspectives or sought to challenge them. Each SWOT theme was addressed separately: first Strengths, then Weaknesses, and so on. But discussions that fell into more than one category were allowed (e.g., an idea to transform a weakness into an opportunity).

A museum staff member worked as an auxiliary facilitator during the sessions. Both author and auxiliary facilitator directed the discussion. Having two facilitators can help group discussion flow better (Rio-Roberts, 2011). The protocol and some more information about the SWOT focus groups appears in Appendix A.

3.3.3 Surveys

Self-reporting can be a complex task for a respondent (Dunning et al., 2004), and several forms of bias can affect the validity of the results. Potential biases include socially desirable responding, acquiescent responding, and extreme responding (Paulhus & Vazire, 2007), identity related biases (Brenner & DeLamater, 2016) and jokesters (intentional false responses) (Fan et al., 2006).

However, self-reporting remains the most commonly used assessment mode (Dunning et al., 2004; National Research Council, 2009), because, when appropriately conducted, a survey can effectively provide quantitative descriptions of trends, attitudes, and opinions of a population (Creswell, 2009; Dunning et al., 2004; Hooper-Greenhill, 2002; Paulhus & Vazire, 2007).

It is often assumed that honesty is enough to produce accurate self-reporting (Paulhus & Vazire, 2007), but the 'illusory truth effect' makes people genuinely believe their knowledge or understanding is correct if they feel familiar with the topic—irrespective of whether it is right or not (Mbewe et al., 2010; W.-C. Wang et al., 2016). Studies using self-reports can employ triangulation with other sources of data to increase confidence in their findings (Paulhus & Vazire, 2007).

Survey length

Visitors may be willing to participate in a survey, but they also have time limits; they are interrupting their own agenda to help another's (Falk & Dierking, 2016; Jacobsen, 2016; Vaske, Jacobs, Sijtsma, & Beaman, 2011). Keeping the survey short is an important acknowledgement of visitors' time constraints, and minimizes the risk of a respondent's attention digressing due to tiredness (Adigüzel & Wedel, 2008). Short surveys are especially important with children, as they may have shorter attention spans (Boyden & Ennew, 1997).

iPad surveying

Surveying on paper is simple, but it also has disadvantages, such as requiring printing, organizing, manual data capture, and physical secure storage for years (King et al., 2013). Electronic devices have overcome or reduced these drawbacks, and can be time-saving for response to closed questions (Davis, Thompson, & Schweizer, 2012; Giduthuri et al., 2014; Leisher, 2014). Using tablets for data collection also allows for the possibility to randomise question order, and present questions in a more visually uncluttered manner (Fowler, 2013).

Compared to paper, electronic surveying produces equivalent results in terms of missing data, item means, and internal consistencies (Davis et al., 2012; Giduthuri et al., 2014; Gwaltney, Shields, & Shiffman, 2008; King et al., 2013; Ravert, Gomez-Scott, & Donnellan, 2015), response rates (Hohwü et al., 2013; Ravert et al., 2015; Shah et al., 2016), and time spent completing the survey (Shah et al., 2016).

Tablet and PC-based surveying present comparable data quality (Struminskaya, Weyandt, & Bosnjak, 2015; Wells, Bailey, & Link, 2013), but paper-sized tablets have the advantage of being attractive for face-to-face surveys (Leisher, 2014). They are not only completed more quickly, but better-ranked in ease of use, efficiency, satisfaction, and accuracy of answers than PC/Web-based questionnaires (Sasao, Konomi, Arikawa, & Fujita, 2014).

Electronic devices have been successfully used with children (Kano & Read, 2012), young people (J. L. Jones & Sinclair, 2011; Ravert et al., 2015), old people (Gwaltney et al., 2008; Shah et al., 2016), and even rheumatic patients (Richter et al., 2008). Although sometimes younger visitors need to instruct their older relatives on their use, the appeal to use iPads in an informal setting is independent of age, gender and group composition (Serrell, 2015).

Surveys for this research were created on SurveyGizmo™. A pilot was conducted with paper and iPad versions. After confirming there was no evident difference, paper surveys were only kept for emergency (e.g., internet down) or visitors that preferred paper. The latter scenario didn't happen, not even for an old man with light Parkinson's or an old woman who didn't bring her reading glasses. Both declined the offer to swap the iPad for the paper version. The man was able to fill the survey out. The woman found an easy to her problem by the zooming in with two fingers. Figure 3-1 shows a picture during the data collection at Tūhura.

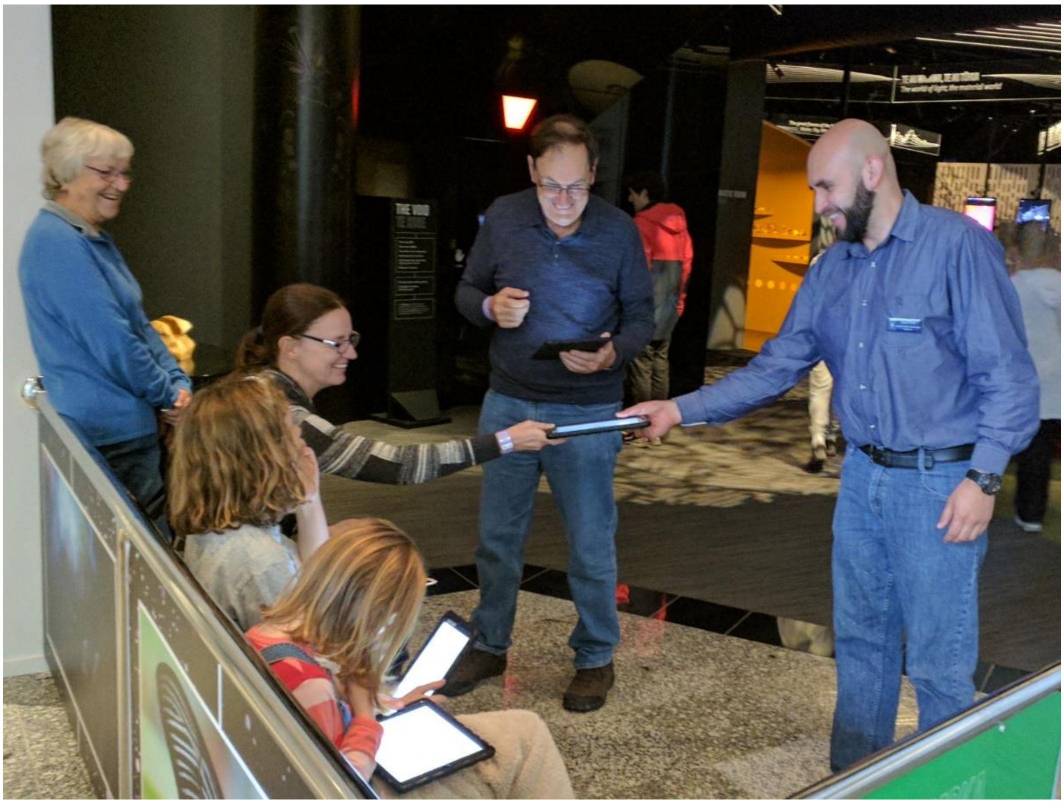


Figure 3-1. The author (right) collecting data on iPads from a three-generation family at Tūhura. Photography taken and reproduced with permission from the participant family.

Non-monetary incentive

Comparatively low response rates can produce similar results to those with high response rate (Keeter, Kennedy, Dimock, Best, & Craighill, 2006; Kohut, Keeter, Doherty, Dimock, & Christian, 2012), but increasing the response rate improves the validity and quality of survey data. This can be achieved by giving respondents an incentive (Coughlin et al., 2011; Edwards et al., 2009; Fowler, 2013; J. M. Griffin et al., 2011; Guo, Kopec, Cibere, Li, & Goldsmith, 2016; Laguilles, Williams, & Saunders, 2011; Olsen, Abelsen, & Olsen, 2012; van Gelder et al., 2018; Wilson, Petticrew, Calnan, & Nazareth, 2010).The incentive used in this study was a small token (Figure 3-2) consisting of a glow in the dark item or magnetic butterfly. The token was attached to a piece of paper with a scientific fact and it was given after completing the post-survey.

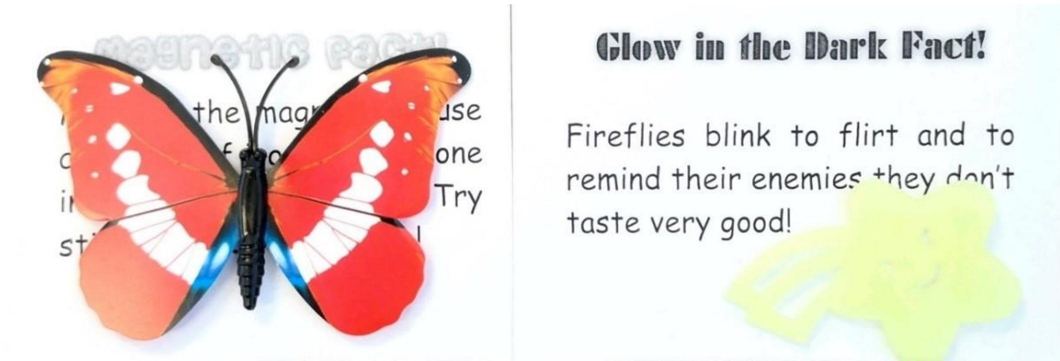


Figure 3-2. Two examples of non-monetary incentives for survey respondents.

3.3.4 Interviews participants

Otago Museum's Decision Makers and Science Communicators (floor staff) were interviewed to explore their views with respect to science, science communication and the redevelopment (see Chapter 2). To be eligible, they had to be involved with either the redevelopment and/or the running of the science centre, and be working at least 0.8 full-time equivalent. All of the Science Communicators that fulfilled these conditions were interviewed. Most of the Decision Makers who fulfilled them were interviewed as well. Only a couple of Decision Makers who were not available and were not interviewed.

3.3.5 Focus groups participants

Focus groups were aimed specifically at solving the research question four (Table 1-6) about what exhibit characteristics influence science learning. Achieving such a complex enterprise required participants not only committed to spend hours in a highly demanding activity in terms of mental focus, but capable of deep insights. Participants should also have a good level of science knowledge, so that they can provide valuable insights due to their knowledge of the phenomenon of interest (Hernández et al., 2014; Kitayama & Cohen, 2010; Palinkas et al., 2015). It's unlikely these requirements would have been met by random visitors and so they were not considered as viable participants. While these characteristics might be fulfilled by staff, they are too closely involved with the exhibits and that might stop them from objectively criticizing them. Moreover, their knowledge about science and understanding of the exhibits' functioning is higher than an average visitor, making them not see the exhibit as a normal visitor would.

Since participants were expected to have the fresh point of view of the occasional visitors that interacts with the exhibits with no prejudices of knowledge, while still having a good understanding of science, commitment and capacity of deep insight, expert sampling (a.k.a. purposive samples, judgmental sampling), was deemed more appropriate than random selection. Expert sampling selects participants by applying expert knowledge of the population (Battaglia, 2008; Etikan & Bala, 2017; Etikan, Musa, & Alkassim, 2016; Hernández et al., 2014; Palinkas et al., 2015). There are many variations of expert sampling. In this case, "[i]ndividuals are selected based on the assumption that they possess knowledge and experience with the phenomenon of interest and thus will be able to provide information that is both detailed (depth) and generalizable (breadth)" (Palinkas et al., 2015, p. 539). Expert in this context does not mean a scholar, but a person deeply immersed and competent in the culture from which the participants are drawn (Kitayama & Cohen, 2010). The decision was to work with children from eight to 18 years old. The main two reasons were availability and that in that age range it was possible to find children engaged with science while not yet focused in science or other field (as it happens during the university studies). They also would bring the fresh perspective that was expected to work better with the method. A museum staff member[23] with many years of experience as a science educator and communicator was the perfect expert to select the participants. From his vast pool of child acquaintances, he identified participants who he considered smart and committed for two focus groups[24]. One group was composed of children between 8 and 12 years old, and the other of adolescents between 13 and 17 years old[25].

3.3.6 Survey participants

Target population

According to Piaget's theory of cognitive development (Piaget, 1968), children enter the 'concrete operational stage' at around seven or eight years. Since the survey questions require a certain maturity to answer correctly, respondents were required to be at least eight

[23] The same one that worked as facilitator.
[24] McKenna-Cress and Kamien (2013) recommend participants to don't know each other. However, a pilot focus group was conducted following this recommendation and where children and adolescents participated together. Two things were evident: first, the group should be split in younger children and adolescents. Second, participants not knowing each other is not the right approach with children, as they felt intimidated to speak in front of unknown children. The final selection of participants by the expert was done under these considerations.
[25] The two sessions were conducted one year apart. Age ranges presented are for the first session.

years old. “By eight years of age children have already amassed a great deal of information” (Lindon, 1996, p. 50). Surveying young children is adequate, they are able to respond questionnaire from seven years old (National Research Council, 2009).

Obtaining a representative sample of groups with different levels of mental maturity or worldviews was expected. The first group was named ‘Children’ (8-12 years old). As explained in Section 1.3.3, a division between adolescence and adulthood is also convenient. The second group, ‘Adolescents’, was defined as from 13 to 18 years old. The decision of taking 18 years old is based in several things. “In secondary school, the 11- to 18-year-olds may progress to very complex levels of dealing with abstract ideas and symbols, especially in mathematics and science” (Lindon, 1996, p. 53). Secondary school in New Zealand is typically attended from 13 to 18 years old (New Zealand Education, 2019). The Otago Museum sells child tickets to up to 18 years old visitors (Otago Museum, 2019).

Alongside age and gender, prior knowledge and context can influence learning (see Section 1.3). Including all possible prior knowledge levels and contexts is beyond the scope of this research. However, it is still possible to acknowledge the influence of these variables in terms of cohort effects.

> Some of the differences that can be seen across age groups do not disappear as individuals age. Instead, they serve as distinctive characteristics of a particular generation. These are known as cohort effects, meaning that they are attitudes, traits, or behaviours that typify a group of people born during a specific period, and they tend to stay with that cohort consistently across the life course. (Fenichel & Schweingruber, 2010, p. 155)

Andert (2011) reviewed the characteristics of some generations. The Baby Boomer Generation (1943-1960) is predisposed to working synergistically. Generation X (1960-1980) is technologically savvy, but shares similar values and characteristics with the Baby Boomers—so much so that it isn’t always a separate group, but the concluding stage of the previous generation. The Millennials (1980-2000) grew up with computers and seek learning opportunities and challenges.

Even though the delineation of cohorts is somewhat arbitrary, it allows generations to be to differentiated, particularly in terms of experiences with and attitudes toward technology (Fenichel & Schweingruber, 2010). Baby Boomers had TV and satellites, but no personal computers until well into adulthood. Generation Xers experienced the computer revolution when they were teenagers. Millennials not only grew up with personal computers, but also experienced a full range of technological tools, including internet and social networking (Fenichel & Schweingruber, 2010).

Given the above, it was decided to divide adults into 'Young Adults' (19-40) and 'Mature Adults' (41+). Young Adults is mainly composed of Millennials. They were expected to be digital natives with either no children or young children[26]. Mature Adults contains mainly Baby Boomers and Generation Xers. It was expected they were not digital natives (although they can be tech-savvy) and their children (if any) may already be in adolescence or adulthood. Some Mature Adults were expected to visit the centre with grandchildren.

Data collection limitations

The closure of Discovery World limited the amount of data collected from it. Since sampling immediately after a major change is not advisable (Frechtling et al., 2010), surveying at Tūhura was postponed five months after opening. Tūhura's data collection period happened one year after Discovery World's.

Also, although a sample can mirror the population if the right measures are taken (Fowler, 2013), it is not expected that visitors will represent the general population. Visitors are self-selected from the moment they decide to go to a science centre (Farrell & Medvedeva, 2010), especially considering such a visit is not free. In 2019, Tūhura's entrance fee was NZD$15 for adults and NZD$10 for children (Otago Museum, 2019). This is similar to the members of the Association of Science-Technology Centres (ASCT), whose median admission charge was of US$10.75 for adults and US$8.85 for children in 2016 (Association of Science-Technology Centers, 2017).

Sampling

Random sampling (a.k.a. probability sampling) is used to minimize systematic variations in participant selection (Creswell & Clark, 2017; Etikan & Bala, 2017; Field, 2013; Hernández et al., 2014; Kitayama & Cohen, 2010). It is a widely accepted method (Creswell, 2009) where randomization can be obtained in several ways. For example, asking every fifth visitor.

Nevertheless, unless the acceptance rate is 100%, any sample based on volunteering violates strict randomness; those responding to the survey can only be said to be representative of volunteers (Couper, 2007; Fowler, 2013; Yeager et al., 2011), not the whole population. Still, random selection is preferable to self-selection, as the latter would alter representativeness even more.

[26] The mean age for the first-time mothers in New Zealand is roughly 28 years (Statistics New Zealand, 2012).

Given that "there is no single answer as to how visitors or professional audiences should be sampled" (A. J. Friedman, 2008, p. 42), this study proceeded on the premise that visitors are not featureless numbers. Therefore, common sense should prevail over cold randomness and the research should not pose any collateral discomfort to visitors. All visitors coming through were asked to complete the survey, except for the following conditions[27]:

- Only one iPad[28] was available.
- The number of children under eight was smaller than the number of possible respondents in the group, minus two[29].
- They had an impediment.[30]

3.3.7 Sample size

There was no need to calculate a sample size for the number of interviews, as almost all of the staff was involved. A focus group should have a size so that it generates enough discussion, but not so large it becomes difficult to manage (Barbour, 2008; Rio-Roberts, 2011). Recommendations vary quite a bit, like ten (McKenna-Cress & Kamien, 2013), six to twelve (Rio-Roberts, 2011), seven to ten (Hernández et al., 2014; L. E. Sullivan, 2009) or six to eight (Fowler, 2013; Hennink, 2013). What really matters is that the group be "small enough for everyone to have an opportunity to share insights and yet large enough to provide diversity of perceptions" (Krueger & Casey, 2000, p. 10). It is perfectly possible to have a focus group discussion with as few as three or four participants (Barbour, 2008).

Lastly, when it comes to surveys, a common belief is the more, the better. But the cost of larger sample sizes may exceed the benefit. Larger samples reduce sampling error, but do not reduce response bias and do not always provide a better study (Frechtling et al., 2010; Naing, Winn, & Rusli, 2006).

Daniel and Cross (2018) provide a sample size formula with finite population correction. This formula saturates at N~384, and that figure became the goal for this research. Due to time constraints, only two surveys in Tūhura surpassed the threshold (N=464, N=386). However, the other one from Tūhura (N=249) and the one from Discovery World (N=224)

---

[27] The protocol is more extensive; this is the summarized version.
[28] Surveys were filled out on iPad.
[29] Only eight years and older were allowed to fill out the survey, but non-respondent children were not allowed to go into the science centre without adult supervision (science centre policy). During the pilot, it was seen that these conditions were causing discomfort to younger children, as they were not doing anything. This exception rule relaxed things and, typically, one or two in the group didn't answer the questionnaire to entertain the kids. This behaviour was not suggested; it naturally happened. But when all respondents decided to participate, the author used his smartphone to show an interactive animation to the kids to help them pass the time.
[30] For example, a person with mental paralysis and their caretaker were not asked; in another circumstance, someone using crutches had no space to sit on the couch, and therefore was not asked.

still had acceptable sizes, as there is actually only a modest gain in precision when increasing the sample size above 200 (Fowler, 2013).

The response rate in Discovery World (87%) was determined from the number of rejections (38) and the number of pre-surveys completed (262). However, since visitors don’t come individually, but in groups, using the previous form to calculate the response rate may inflate the result. Even though it is a common practice to report response rate, other research that implies asking groups often does not explicitly state how it derives said rate. Thus, it is proposed here to divide the number of groups that were asked (582 in Tūhura) by the number of groups that accepted (456). Tūhura’s rate was 78%.There is no agreed standard for a minimum acceptable response rate (Fowler, 2013), but surveys that achieve a rate of 70% or higher are generally considered high-quality and nonresponse is not a concern (Newcomer & Triplett, 2015). See Appendix E for more information on sample sizes and response rates.

3.3.8 Demographics

Interviews and SWOT focus groups

Seven interviewees were male and seven female. Seven were Science Communicators, five Decision Makers, one a Māori curator and one a technician. No other demographics were collected.

Two focus groups were held at Discovery World, one with younger children and one with adolescents. The intention was to repeat the focus groups with the same participants after the redeveloped science centre opened. Ten children between 8 and 12 years old (6 male and 4 female) and eight adolescents between 13 and 17 years old[31] (7 males and 1 female) participated in the Discovery World focus groups. For the Tūhura groups, several previous participants could not continue—they were either busy or out of town. No replacements were added so as to not introduce noise by adding participants without a Discovery World experience. Fortunately, those who were able to come were the most active members of the previous groups. Four children (1 male, 3 female) and five adolescents (all male) participated in the Tūhura focus groups. It is acknowledged the potential bias of the all male adolescents focus group.

Surveys

Surveys were collected at Discovery World in June and July 2017, and at Tūhura from May to October 2018. Demographics from these surveys can be seen in Table 3-1.

[31] At the time of the first focus group, for Tūhura’s focus groups, they were 1 year older.

Table 3-1

Demographics from surveys at the science centre (Discovery World / Tūhura) and the whole museum (In-house survey / MoA survey)

		Discovery World	Tūhura	In-house survey	MoA survey
Gender (N_{DW}=224/ N_T=1097)	Female	62	59	64	
Age (N_{DW}=224/ N_T=1084)	8 to 12	29	27		
	13 to 18	16	19		
	19 to 40	35	30		
	41 plus	35	24		
Visits (N_{DW}=224/ N_T=963)	Recurrent	52	33 (55)	61	65
Company (N_{DW}=224/ N_T=1034)	Family	77	75		
Ethnicity (N_{DW}=224/ N_T=1067)	European	76	87	47	
	Māori/Pacific	12	12	2	11
	Other	12	10	51	
Residency (N_{DW}=224/ N_T=1068)	Otago	50	53	59	49
	Overseas	9	8	32	26

Recurrent visitors to Tūhura were calculated twice, not considering visits to Discovery World, and considering them (in brackets).
In Discovery World's survey, ethnicity was a closed question where it was not possible to select more than one. But then it was clear the New Zealand has a high number of citizens with more than one and Tūhura respondents were allowed to choose more than one ethnicity.

For comparison, the table includes (when available) demographics from two other surveys conducted at the Otago Museum, one internal (In-house survey, December 2014) and one under the methodology of Museums of Aotearoa (MoA survey, March 2015) (Otago Museum, 2015). These data are for the whole museum, not only the science centre, and both were conducted with adult respondents.

Notice the high influx of female visitors. According to Serrell (2015) the percent of males and females at museums are not usually different, but Ecsite-uk (2008) reported that in 10 of the 14 organizations they asked, females outnumbered males, with a total average of 56% of visitors being female.

Visual counting

As discussed above, survey demographics are not the demographics of the whole population, but visually counting and assessing visitor demographics can be used to quantify visitation and compare exhibitions (Randi Korn & Associates Inc., 2006; Sandifer, 1997).

Discovery World and Tūhura visitors' demographics were unobtrusively assessed[32] and recorded by the author during the surveying period.

Table 3-2 provides comparative demographics from visitors to Discovery World and Tūhura with the New Zealand Census 2013[33] data for the Otago Region (Statistics New Zealand, 2019). Since age groups from the census do not coincide with the visual assessment groups, linear interpolations were made to calculate the relevant ranges.

Table 3-2

Demographics from visual assessment at the science centres and from the New Zealand Census 2013 in the Otago Region

		Discovery World	**Tūhura**	**NZ Census**
Gender	Male	43	44	49
(N_{DW}=1497 / N_T=3301)	Female	57	56	51
	< 2	8	6	2
Age	2 to 7	20	26	6
(N_{DW}=1625 / N_T=3493)	8 to 12	11 (15)	12 (18)	5 (5)
	13 to 18	7 (10)	8 (12)	9 (10)
	19 to 40	31 (44)	30 (44)	30 (33)
	41 plus	22 (31)	18 (26)	48 (52)

Since gender in toddlers under two years old was not assessed, the sample size of Gender is smaller than the sample size of Age.
To facilitate comparability with survey demographics, the equivalent demographic percentages of not considering people under eight years old are shown in brackets.

Most visitors to both Discovery World and Tūhura came in groups of two generations (76% and 78%, respectively). From these two-generation groups, most included children[34] (88% in Discovery World and 89% in Tūhura). From the two-generation groups with children, two thirds came in small groups of two to four members (67% in Discovery World, 69% in Tūhura).

Not surprisingly, the proportion of young children attending the science centre is a lot higher than the general population's proportion. For both science centres, the proportion of female attendees (56%, n=2708, CI=1%) is significantly different (difference=5%, CI(4%,7%)[35], $\chi^2(1)=50.5$, $p<.001$) than the female proportion in the Otago Region (51%, n=103.9k, CI=0.2%). Higher female attendance has been previously detected in the museum

[32] Gender was not assessed for children under 2 years old.
[33] The 2018 Census has not yet been completely released. It was announced that would happen in 2020.
[34] Most likely they were families, as school groups were not included.
[35] CI is reported as a range instead of a single figure because the range was not symmetric.

and is not surprising, “as caregivers are more likely to be female” (Otago Museum, 2015, p. 43).

The distribution of visitors is heavily tilted towards adults (Table 3-2). If survey demographics were representative of visitors, we would have ended with very few children’s responses. However, the only bias observed in this sampling method is that age groups were more evenly distributed. It can be considered that the sampling method was pseudo-stratified, i.e., the non-homogeneous population was stratified (in age groups), and sampling was such that it produced more precise estimates of each stratum (Etikan & Bala, 2017). The reason for this to happen is that many science centre visitors come in families composed of adults and young children. Since younger children could not participate, those adults could not either and the overall presence of adults in respondents decreased in favour of older children.

3.3.9 Design of ordinal items

Likert-type items

In an ordinal item, a respondent chooses from a set of ordinal options. The most popular is the Likert format (Likert, 1932), developed almost a century ago. In this form, choices vary from ‘strongly disagree’ to ‘strongly agree’, centred around a neutral midpoint (‘neither’). Ordinal items are a popular way to collect self-reports and there are multiple variations around the Likert format.

Although sometimes the term ‘Likert-type scale’ is used for single items in Likert form (Adelson & McCoach, 2010; Lantz, 2013), this is misleading. Single items are not scales. A better terminology is ‘Likert-type items’ (Boone & Boone, 2012). The Likert format, even simple in appearance, has several characteristics that can affect its performance. The following are a few that deserve discussion prior to any data collection.

Reversed items

Reversed (negatively worded) Likert-type items are a common approach to avoid response bias. The theory is that a thoughtful respondent should detect the inversion and answer accordingly. Failing to do so would indicate inattentive respondents. But, this is not exactly so. There is no difference in reliability between negatively and positively worded questions, but only when they are considered separately (Borgers et al., 2004). Mixing questions may lower reliability for two reasons: positivity and negativity are not symmetrical[36], and the inversion may be unclear to respondents (Alexandrov, 2010; Van

[36] i.e. a reversed item cannot simply be reversed again to make it positive and compare it with the rest, as the strength of a negative statement and a positive statement are not the same.

Sonderen, Sanderman, & Coyne, 2013).Given the above and the age of the respondents, it was decided to avoid mixing reversed items in the scales.

Middle point

The original Likert format is symmetrical around a neutral point (Likert, 1932). Some researchers do not consider it a true neutral, as it is often used by undecided respondents (Raaijmakers et al., 2000) or unmotivated respondents (Masuda et al., 2017). In theory, by not having this option, undecided respondents would be forced to make up their minds and choose a 'real' option. Under this logic, some researchers have suggested that removing the middle point can improve the reliability of the instrument (Baeza, Tort, Romá, & Benito, 2001; Borgers et al., 2004; Cardiel & Pattison, 2014; Hernández et al., 2014; Nadler et al., 2015).

Despite this, the possibility of a midpoint having a double meaning (true neutral and indecision) is not an argument to withdraw it (Raaijmakers et al., 2000). Also, there is evidence that scales with an odd number of options outperform those with an even number (Adelson & McCoach, 2010; Borgers et al., 2004).

Removing the midpoint would bias the true neutral responses by forcing them to choose a different option, likely on the positive side (Worcester & Burns, 1975). And there is no evidence that undecided responders are, in fact, making up their minds, or they are still undecided and just choosing a different option. Removing the midpoint is sweeping the underlying problem of indecision under the carpet. Therefore, the option was kept.

I Don't Know (IDK) option

Even though keeping the middle point avoids adding bias, it doesn't solve the problem of undecided respondents. One alternative is adding a nominal "Don't know" (or "No opinion") for those who legitimately don't know or are unwilling to do the mental work required by answering (Fowler, 2013; Krosnick et al., 2002; Pearce-Morris et al., 2014; Raaijmakers et al., 2000).

I Don't Know (IDK) was added to all ordinal items[37] and IDK responses were treated as missing data (Fowler, 2013; Rennie & Williams, 2002). The instrument with highest percentage of IDK responses was the self-efficacy instrument. Before the visit it was 10%. After the visit it was 6%. A plausible explanation is that some respondents unfamiliar with an asked concept selected IDK instead of 'Not confident at all', increasing the IDK rate. In

[37] Except the exploratory visual discrete-like item.

all other scales, the maximum didn't surpass 3%.IDK responses can be interpreted as uncertainty due to lack or partial knowledge. But they can also mean metacognitive awareness, where visitors know that they don't know; this is a sign of critical thinking (Thuneberg & Salmi, 2018). In any case, it is an important option to include (Raaijmakers et al., 2000; Thuneberg & Salmi, 2018), as it reduces guessing and increases reliability (Muijtjens, Mameren, Hoogenboom, Evers, & Vleuten, 1999).

Number of options

The number of options in an item needs to be enough to allow appropriate discrimination, but not so many that it becomes difficult to discern the difference between two adjacent points. Maximum scale reliability is obtained when items have six or seven options (Borgers et al., 2004). However, participants with low discriminant capacity (such as young children) may require fewer categories, such as three (Bourke & Frampton, 1992; Hernández et al., 2014; J. Jacoby & Matell, 1971).

Due to the young age of some respondents in this research (as young as eight years old), ordinal items in the Discovery World survey had only 3 ordinal choices[38] (plus IDK). But, since results showed that even the youngest respondents could manage more than three choices (full discussion in Chapter 6), ordinal items in Tūhura's surveys were created with five options (plus IDK).

Data collection with single items

Whether a single item can be valid to collect data is a matter of discussion. To Rossiter (2011), only semantic differential—a related, but different, format—can work as single items without further issues. However, Alexandrov (2010) concluded that Likert-type single items work equally as well. Rossiter and others consider that Likert-type items can be valid and reliable under some circumstances (Bergkvist, 2014; Rossiter, 2011; Waddimba et al., 2016). The measured construct should be doubly concrete (have a simple, clear object and single-meaning attribute) and be unidimensional (Fuchs & Diamantopoulos, 2009). Despite the controversy and theoretical limitations, the use of individual Likert-type items is common (Clason & Dormody, 1994; Pekarik, Schreiber, & Visscher, 2018).

Icon-based anchors

Given that different wording of the same questions can produce different replies

[38] Except the exploratory visual discrete-like item, that had five.

(Stocklmayer & Bryant, 2012), it was decided to use the same survey for children and adults. Nevertheless, children process information more slowly and need clearer instructions (National Research Council, 2009).

A popular addition to ordinal scales is smiley faces. Children tend to prefer them over fully text-based scales, as they help them interpret the scale (Hall et al., 2016; Read et al., 2002; Reynolds-Keefer et al., 2009). Stange, Barry, Smyth, and Olson (2018) used eye-tracking to discover that respondents to questions with smiley faces spent less time processing the questions and response options than those answering the version with text-only responses. They also found support for respondents with lower literacy relying more on smiley faces than those with higher literacy.

The selection of smiley faces cannot be arbitrary. Chambers and Craig (1998) noticed that ratings in a picture-based scale for pain varied depending on the types of faces used in the scale. The fact is that emoji were born in the era of internet communication to add emotional tone to text (Danesi, 2016). The emotion related to the facial expressions used in icons depends on the knowledge of the code (Danesi, 2016) and ethnic and cultural factors (Danesi, 2016; Reynolds-Keefer & Johnson, 2011). Only a few icons are truly universal (Serrell, 2015). Smiling faces indicates pleasure in most cultures (Ham, 2016).

For instance, the Elementary Reading Attitude Survey uses Garfield as response icons (Kear, Coffman, McKenna, & Ambrosio, 2000). The Disagree option is represented by an angry Garfield. But Garfield is a grumpy and funny character. This may cause participants to find 'disagreement Garfield' an appealing and appropriate option, according to how this character usually behaves. They then may select it despite it not representing their response to the statement (Reynolds-Keefer & Johnson, 2011).

A more typical set is formed by smiley faces running from sad/angry to fun/happy faces. The sad-to-happy range is a good option to rate enjoyment (Longnecker et al., 2014). But sad and angry faces were created to convey sadness and anger, not disagreement. Including them to be interpreted differently from their use in daily life may result in bias. Children hoping to have an enjoyable experience may tend to select only positive ratings (Hall et al., 2016). If children are not, or don't want to be sad or angry, they may reject choosing 'disagreement' (represented by sad and angry faces), biasing the responses towards agreement.

A new set of visual scale illustrations for the Likert scale was designed and drawn by the author (Table 3-3, left-hand side). It is expected that this set outperforms the sad-to-happy and angry-to-happy sets. It is designed to express disagreement ('boo face'), rather than anger or sadness. The faces vary the structure of the mouth and the position of the eyebrows continuously to facilitate the idea of a continuum. Thumbs-up is a concrete and easily interpretable symbol of approval[39] (Danesi, 2016) and so is ‘thumbs down’ as disapproval. Thus, in addition to the faces, ‘thumbs up’ and ‘thumbs down’ were included at the extremes. To keep adding the visual feeling of a continuum with equal distance between consecutive anchors[40], the thumbs turn from down to up at 45° each time.

Table 3-3

Self-concept in science items in Likert format (left) and Visual Discrete format (right)

Likert scale (LS)	**Visual Discrete Scale (VDS)**
“Select the one option for each statement that best shows what you think”	“Click on the penguin that best represents yourself in…”
	
LS1 I have a good understanding of science.	VDS1 …science understanding.
LS2 I could explain some science examples to my friends.	VDS2 …confidence to explain some science examples to your friends.
LS3 I learn science fast.	VDS3 …learning science fast.
LS4 I am good at solving math problems.	VDS4 …ability to solve math problems.
LS5 I am good at solving science problems that do not need math.	VDS5 …ability to solve science problems that do not need math.
LS6 I can understand new science ideas.	VDS6 …confidence to understand new science ideas.
LS7 I usually do well in science.	VDS7 …doing well in science.

3.3.10 Likert-type scale: proved but improvable

<u>Likert format</u>

Once the concept of Likert-type item is understood, the concept of a Likert-type scale comes naturally. It is simply a collection of single items where all of them are in the same format and measure the same underlying construct. Notice that Likert-type scale is used in this thesis instead of Likert scale. This is to acknowledge that this kind of scale has evolved and it does not necessarily keep the original Likert’s format in all the cases.

[39] This set of emoji was created thinking of the western culture only. Thumbs-up, for instance, is offensive in the Middle East (Danesi, 2016).

[40] It is debatable whether Likert-type scales can be analysed parametrically (see Appendix E). This set of emoji may help towards the possibility of interpreting Likert-type scales as interval, but this still needs to be proved.

Numerical values are then assigned to each option in each ordinal item, so that the result can be interpreted as interval (see further discussion on the topic in Appendix E), rather than ordinal (e.g., strongly disagree = 1, disagree = 2, …, strongly agree = 5). The score of a scale is technically the sum of all item scores, but for ease of comparison, it is common to normalize the scores by dividing them by the number of items, so that the new score range coincides with that of individual items.

The problem of verbal labels

Questionnaires are not only methods to elicit information, but a source of information that respondents use to determine their answer as well (Schwarz, 1999). "People respond to the ordinal position of categories as well as to the descriptors" (Fowler, 2013, p. 89)

Likert-type scales (Likert, 1932) are commonly used to assess behavioural constructs, but they can be problematic (McLeod, Pippin, & Wong, 2011). "Usually, researchers will have more reliable, valid, and interpretable data if they avoid the agree-disagree question form" (Fowler, 2013, p. 91).

One issue with Likert-type scales is that scale points are connected to verbal anchors (response labels). Since words don't have a natural and predefined order, they make scale points appear non-equidistant (Lantz, 2013). In other words, is the 'distance' between Strongly Disagree and Agree the same as from Neither to Agree? This issue affects the validity of considering the scale as Interval (H. H. Friedman & Amoo, 1999). It has been proposed that the Likert 'fast form', with labels only at the ends, alleviates some of the Likert biases (Friborg, Martinussen, & Rosenvinge, 2006; McLeod et al., 2011). However, familiarity with these labels (which is language and country-dependent) may still bias the results, as the more familiar the respondent is with an endpoint's wording, the more likely they'll choose that option (e.g. "Completely agree" is more familiar than "Strongly agree") (Weijters, Geuens, & Baumgartner, 2013).

Common method bias refers to the variance (both random and systematic) attributable to the measurement method, rather than to the construct. It is one of the main sources of measurement error (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). One of these biases is acquiescence bias—that is, the propensity to respond positively to items irrespective of item contents. Considering that acquiescence bias in Likert-based response formats may be stronger when only positively-worded items are present, some authors have suggested introducing negative statements to reduce the bias (Baumgartner & Steenkamp, 2001; Podsakoff, MacKenzie, & Podsakoff, 2012). However, negatively-worded labels are less effective at helping young children communicate judgments (Hall et al., 2016). They may even induce new errors, due to the counter-intuitive nature of negative constructions (Friborg et al., 2006).

It has been also proposed to reverse items using polar opposites, but avoiding negations (e.g. a little/a lot) (MacKenzie & Podsakoff, 2012). However, when regular and reversed items are combined in the same test, they produce results with more noise and error, threatening reliability and validity (Suárez-Alvarez et al., 2018).

Alternatives to the traditional Likert-type scale that that children find less enticing are the Numbered Ruler-type Scales (Mellor & Moore, 2013), the Visual Analog Scales, and numeric Visual Analog Scale (van Laerhoven, van der Zaag-Loonen, & Derkx, 2004).

Symmetry versus full positivity

Munshi (2014) studied response points in a 7-point Likert-type scale with eight items and found that distances are not equally distributed. For example, the distance from the neutral point to simple agreement was 25% larger than from simple agreement to strong agreement. The author considers that, although negative and positive responses on either side of the neutral points can be seen as symmetrical (Munshi, 2014), having only positive responses may help toward perceiving responses as equally distributed. This is based on the assumption that visualizing only positive integers is easier than visualizing positive and negatives and zero. Positive integers are easy to understand; zero and negative numbers less-so (De Cruz, 2006). "At first sight, negatives may seem counterintuitive because they violate ontological expectations" (De Cruz, 2006, p. 317). Although zero may seem less counterintuitive, the full use of zero was discovered only by two civilizations, Indian and Mayan (Joseph, 2008).

3.3.11 Visual discrete format as an alternative to Likert format

An alternative to the traditional Likert format was created that transfers the weight from verbal labels to visual labels. Preferably, there would be no need of words. There is little work done on visual alternatives to the Likert-type scale, but Reynolds-Keefer et al. (2009) found no variability in responses of young children when comparing three pictorial Likert-type scales. One used words (capital and small letters for words NO, no, yes, YES) and two emoji (traditional smiley faces and sun-smiley faces) varying from angry to happy.

Images were decided as anchors because they are more familiar and require less cognitive effort to process than their equivalent verbal stimuli (Hirschman, 1986). They need to be chosen carefully, as the features of an image influence the sentiment (Siersdorfer, Minack, Deng, & Hare, 2010). Cuteness is especially important in the digital culture (Wittkower, 2012).

Pilot: Visual discrete-format item

A bunny was used as anchor in Discovery World's pilot. But since rabbits are an invasive species that are considered a problem in New Zealand (Norbury & Jones, 2015), it was swapped for a penguin, an endemic bird, in Tūhura. The pilot was conducted with a single item. The item had five points, no IDK option, and still included verbal labels under the images.

Final version: Visual Discrete Scale

The format was well received in the pilot, and no major problems were evident. The final version achieved the goal of eliminating labels. The newly developed Visual Discrete Scale can be seen in Table 3-2 (right-hand side). The format has the following characteristics:

- Visual anchors. Language bias is eliminated by not including labels.
- Cuteness. The icon was especially chosen to be cute, so that people were more enticed to answer the questionnaire.
- Homogeneity. By having the same image in all the options, the bias of one image more enticing than other is avoided.
- Proportionality. The size of the image varies in size proportionally to the level of agreement.
- Positivity. All the options were positive, eliminating the possibility of a bias coming from comparing the negative and positive sides of the format. Respondents that in Likert format choose the midpoint can still do so (level three), but unsure respondents are not drawn to a midpoint as it is no longer a 'Neither'.
- IDK. Unsure respondents still have the option of I Don't Know.

3.3.12 Assessment of self-beliefs in science

Since Discovery World was due to close a few months after the research started, there was no time to do multiple pilots. The decision was to include several constructs that might[41] fulfil the conditions necessary to be assessed with a single item. From the pilot's results, the more promising self-beliefs were assessed thoroughly with full scales in Tūhura.

[41] This was not proved at the moment, but it was done in Tūhura.

Appreciation of science and science outreach

Two single Likert-type items in Discovery World's survey assessed constructs related to appreciation[42]: L1. I think science museums are important (appreciation of science outreach), L2. Scientists' work is important to me (appreciation of science). Both had three ordinal options: Disagree, Neither agree nor disagree, Agree, plus a nominal I Don't Know.

Self-concept

Discovery World gave the opportunity to explore the construct of self-concept in science with two exploratory single items. The first one was presented in three-point (plus IDK) Likert-type format: L3. I could explain some science examples to my friends.

To test the new Visual Discrete format, self-concept in science was also assessed in Discovery World with this format. The item was 'Click on the penguin that best represents yourself in science understanding'. It had five ordinal options (no IDK) accompanied by small text labels: Brand new, Beginner, Capable, Skilled, and Expert. Self-concept was chosen as the construct to be assessed while testing the new format in full, because it was theorized that its stability would help minimize the effect of unknown, uncontrolled variables (lurking variables). From the results, it was determined that IDK is a desirable option and children could manage 5 ordinal choices, as suggested by Adelson & McCoach (2010). Given the promising results, labels were completely removed and a full scale was developed.

Even though it is not a common method, starting out with a single exploratory item and later constructing a full scale has been used before. The International Association for the Evaluation of Educational Achievement (IEA), in the first two Trends in International Mathematics and Science Study[43], used 'I usually do well in science' as an isolated item about self-concept in science[44] (IEA, 1994, 1998). From there, it evolved to a full scale (IEA, 2002) that currently has six or eight items, depending on the school grade (IEA, 2018a, 2018b). The original single item was compared to the full scale and determined to represent a substantively valid measure (Wilkins, 2004).

Since the Visual Discrete Scale still had to be tested, a full scale about self-concept in science was developed (described below in Section 5.4) for Tūhura in both Likert-type

---

[42] There was a third item, 'Science is present in my life', but children increased their IDK responses after the visit. Due to lack of reliability it was removed from further analysis (Frechtling et al., 2010).

[43] Seventy countries were expected to participate in 2019, with students in fourth through eighth grade (Mullis & Martin, 2017).

[44] The questionnaire was actually quite comprehensive, with related items, like 'To do well in science you need good luck', but this one is the only one that can truly be identified with Self-concept in Science.

scale format and Visual Discrete Scale format (Table 3-3). Each item is matched one-to-one. The similarity in demographics between the two survey samples (Table 3-4) supports the comparability between their results.

Table 3-4
Percentage demographics comparisons between the Likert scale (LS) and Visual Discrete Scale (VDS) respondents

	Females	**Males**	**8-12**	**13-18**	**19-40**	**41+**
LS	59	40	24	18	34	25
VDS	59	41	26	21	32	21

Due to missing values, sample sizes vary. Gender is based on N=446 for Likert-type scale and N=372 for Visual Discrete Scale, age distribution is based on N=441 for Likert-type scale and N=369 for Visual Discrete Scale.

Self-efficacy

Self-efficacy is another interesting self-belief that is related to self-concept, but more malleable. Given that all of the statements were related to exhibits at Tūhura's Unseen Forces Zone, 'self-efficacy in light and electromagnetism' was measured, but for simplicity's sake, it will be referred to throughout the thesis as self-efficacy in science. The construct was operationalized with a five-point Likert-type scale (Table 3-5). Labels were modified from the usual agreement level form to a confidence form, from 'Not at all confident' to 'Very confident'.

Table 3-5
Instrument to measure self-efficacy in science

"Please, rate how confident you are that you can…"
SE1. find the position where a prism splits white light into different colours.
SE2. recognize plasma if you see it.
SE3. name characteristics of the infrared light.
SE4. test if a material is affected by a magnetic field.
SE5. recognize whether a colourful room is illuminated with monochromatic light.

From the 227 matched pre-post responses, 28% were Children, 20% Adolescents, 23% Young Adults and 29% Mature Adults. Respondents were mainly female (61%). The instrument is unidimensional (a single factor explained 65% of variance before[45] the visit and 67% after[46]) and the internal reliability is good (α_{pre}=.862, α_{pre}=.874).

Scientific fluency

Scientific fluency (see Section 1.3.3) is a new construct built upon the concept of aesthetic fluency (J. K. Smith, 2014; L. F. Smith & Smith, 2006). L. F. Smith and Smith (2006) developed an instrument to measure aesthetic fluency by asking visitors to rate their level of familiarity with five art terms and five artists. By analogy, scientific fluency was operationalized with an instrument that includes five science concepts relevant to this study (Prism, Plasma, Infrared light, Magnetic field, Monochromatic light) and five scientists (Albert Einstein, Ernest Rutherford, Isaac Newton, Marie Curie, Leonardo da Vinci).

Due to its novelty, the design of this instrument tried to be as faithful to the original as possible. This meant numbering the options from 0 to 4 during data collection (later recoded as 1 to 5 in the analysis) and not including the IDK option. The five response options were: 0 - I have never heard of this (person)[47], 1 - I have heard of this (person) but don't really know anything about it (him / her), 2 - I have a vague idea of what (who) this (person) is, 3 - I understand this when it is discussed (I could understand a discussion about this person[48]), 4 - I can talk intelligently about this (person)).

All of the science concepts included in the construct appeared in at least one exhibit at Tūhura's Unseen Forces Zone and the questions were asked before and after the visit. Since the pre-survey was already considered at its maximum length by the author, the questions about scientists were only included in the post-survey. The exhibition doesn't mention scientists, with the exception of Newton. Given that Fluency about scientists was filled out after the visit, it may be that knowledge about Newton is a bit higher than it was before the visit. However, the inclusion of scientists was not done to assess a change due to the visit, but to keep the format of Aesthetic Fluency as much as possible and explore the possibility of using the full format in later research. Still, it produced some worth noting results.

The sum of the five Familiarity scores for concepts is defined as the fluency in scientific concepts. The equivalent for the people is the fluency about scientists. The sum of both fluencies is scientific fluency. Notice that all of the scientific concepts were related to

[45] Bartlett's test of sphericity $\chi^2(227,10)=510$, p<.001, KMO=.825.
[46] Bartlett's test of sphericity $\chi^2(227,10)=535$, p<.001, KMO=.862.
[47] In brackets appears the change in option from concept to scientist.
[48] The original version said "I understand this artist when it is discussed", but the researcher considered that that structure is fine for artists (artists are identified usually with their work), but "understand a scientist" would be a little out of place.

light and electromagnetism, but for simplicity they are just mentioned as scientific concepts. From the 386 matched pre-post responses, 27% were Children, 21% Adolescents, 31% Young Adults, and 21% Mature Adults. The majority of the respondents were female (59%).

The sub-instrument to measure fluency about scientists was unidimensional (66% of variance explained by a single factor[49]) and highly reliable (α=.867). The sub-instrument to measure fluency in scientific concepts was also unidimensional (a single factor explained 66% of the variance before[50] and 73% after the visit[51]) and highly reliable (α_{pre}=.869, α_{post}=.904).

Despite mixing people with concepts, the full instrument to measure scientific fluency is also unidimensional (58% of the variance is explained by a single construct[52]) and with excellent reliability (α=.917). Unidimensional scientific fluency can be seen in the scree plot (Figure 3-3) from a factor analysis by principal components. This supports L. F. Smith and Smith (2006)'s assumption that familiarity with concepts and people are each part of the same construct.

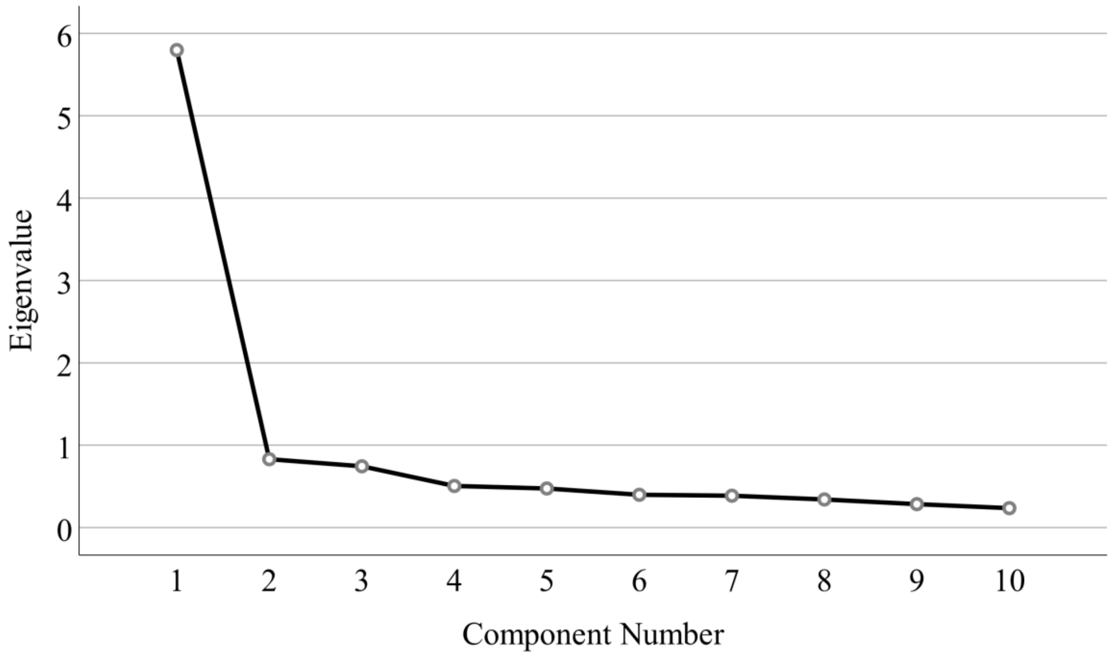


Figure 3-3. Scree plot of scientific fluency. The number of components with eigenvalue over 1 is the number of dimensions (underlying constructs) in the scale.

[49] Bartlett's test of sphericity $\chi^2(386,10)=949$, $p<.001$, KMO=.837.
[50] Bartlett's test $\chi^2(386,10)=884$, $p<0.001$, KMO=0.861.
[51] Bartlett's test $\chi^2(386,10)=889$, $p<0.001$, KMO=0.889.
[52] Bartlett's test $\chi^2(386,45)=2207$, $p<0.001$, KMO=0.927.

3.3.13 Assessment of science engagement

Visiting a science centre is a free choice; visitors go mainly because they (or their family) are engaged or expect to be engaged. Engagement was measured in several ways.

Visit time

Visit time (a.k.a. dwell time) can be used to measure the value an individual assigns to the visit (Jacobsen, 2016). Visit time is linked to engagement and learning (Sandifer, 1997). The same iPad was used for a visitor to complete both pre- and post-survey. iPads were set to record the time between the end of the pre-survey and the beginning of the post-survey. This time coincides with the visit time. Because the Tropical Forest (a contained space with released butterflies and other live animals) is part of the science centre, the total visit time also includes any time visitors spent there.

Engagement scale

In Discovery World, engagement was assessed with a single three-point (plus IDK) Likert-type item: L4. I expect to enjoy Discovery World (pre) / I enjoyed Discovery World (post).For Tūhura, a full instrument was developed. The original intent was to find an engagement scale in the literature and use it. However, with all the scales that were found, only one or two items were a good fit with the definition of engagement in informal settings used in this research. In the end, Tūhura's instrument was formed with individual items from other authors (Table 3-6).

Table 3-6
Engagement with science scale

Item	Reference
E1. I would like to do some scientific research.[a)]	(Füchslin, Schäfer, & Metag, 2018)
E2. I have a real desire to learn science.	(Weinburgh & Steele, 2000)
E3. Science is fun.	(Hodzi, 1992)
E4. I would enjoy being a scientist.	(Hodzi, 1992)
E5. I enjoy learning science.	(Guzey, Harwell, & Moore, 2014)

[a)]The original wording was "I would like to partake in scientific research once" (Füchslin et al., 2018)., but it deemed too complicated and it was modified for younger respondents.

Since during the pilot (n=40) there was no difference between the pre and post-tests (M=4.1 in both), it was decided to only use the scale in the pre-visit survey. The pilot instrument contained another two items. E6. I am curious about the world in which we live.

E7. I would like to belong to a science club (Fraser, 1978, 1981). They were dropped due to their lower correlation with the rest of the scale.

Self-reported Intention and Interaction

Engagement can also be seen through "behaviours that demonstrate interest in or interaction with a science-related activity or experience" (McCallie et al., 2009, p. 20). Detailed visitor observation can provide valuable data on engagement, but it is a time-consuming and expensive method (Frechtling et al., 2010).

Instead, since most visitors arrive with pre-existing expectations for what they will see and do (Falk & Dierking, 2016), they were asked to self-report it. Before the visit, they ticked which of the following they intended to do:[53] Visit the Tropical Forest [a,b], Interact with the exhibits [a,b], See the Science Show [a], Go to the planetarium [b] and Learn some science [b]. After the visit, they were asked to tick off all options they actually did. This method measured how many originally-disengaged visitors became engaged with something at the science centre. Notice that 'Learn some science', and 'Read some panels' are the only two that were not paired. Thus, they are not aimed at measuring the engagement change from before to after the visit, but to work independently as a variable that could potentially explain other results.

From the 224 Discovery World respondents, 29% were Children, 16% Adolescents, 35% Young Adults, and 20% Mature Adults. The main gender present was female (62%). From the 1004 Tūhura visitors that filled out this section, 27% were Children, 19% Adolescents, 31% Young Adults, and 24% Mature Adults. Most were females (59%).

Reporting for themselves and for others

Another question related to engagement was asked after the visit. It read, 'Who do you think Discovery World / Tūhura is suitable for? Tick all that apply'. Visitors were able to tick the options they considered appropriate: Children, Teenagers, and Adults. Notice that, in this case, visitors not only assess how appropriate the science centre is for them, but for what age range(s) they think it is appropriate. From the 437 visitors that answered this question, 26% were Children, 17% Adolescents, 33% Young Adults, and 24% Mature Adults. Fifty-eight percent were female.

[53] Superscript [a] means it was asked in Discovery World. Superscript [b] means it was asked in Tūhura.

Ranking

Ranking options is a way to obtain a visitor's attitudes by measuring how they prioritize the options (Lavrakas, 2008). Tūhura visitors were asked to rank how important (from most to least) Mathematics, Explanations, and Experiments are for learning science before and after the visit. Amongst the 197 visitors that answered the question, 30% were Children, 21% Adolescents, 21% Young Adults, and 28% Mature Adults. Females represented 60%.

Star rating

A simple 5-star rating scale has become a recognisable review system (Ganu, Elhadad, & Marian, 2009; Stoyanov et al., 2015). Tūhura visitors were asked to star-rate Tropical Forest, the exhibits, and the planetarium in terms of fun and learning. Notice how this instrument is related to both learning and engagement.

Due to a software bug, it was not possible to show the specific part of the question to only those who previously selected they visited the corresponding attraction. Therefore, the three attractions were shown to all visitors, asking them to respond only for the attractions that they had visited. No problems appeared with Tropical Forest and the exhibits, but a large number of visitors who did not visit the planetarium still rated it (n=67). It is unclear why they rated something they didn't visit, but given that their ratings were lower than those of the actual visitors, it seems plausible that they didn't know there was a planetarium. Anecdotally, several visitors seemed surprised by the question about the planetarium, asking their companions questions like, 'Is there a planetarium?'.

To improve reliability, responses from visitors that rated an attraction, but previously reported they hadn't visited it, were not included in the calculations. For the same reason of improving reliability, only responses that rated both learning and fun were included.

3.3.14 Assessment of scientific knowledge

One of the staff's main concerns was that Discovery World was seen only as a playground, with little science learning taking place. Rennie and Williams (2002) also reported the consideration that visitors might not learn because they only play with the exhibits or do not read the labels.

There is evidence that a single visit can increase visitor's scientific content knowledge (Falk & Storksdieck, 2005; Fender & Crowley, 2007). But, to the author's best knowledge, no studies have been done where Science Knowledge is assessed objectively with not on a school tour visitors before and after a visit to a science centre (See 1.4.3).

Objective assessments are often not conducted in informal settings to avoid alienating visitors. "Public audiences use exhibitions as a form of leisure experience, not to pass an examination or cover a curriculum" (A. J. Friedman, 2008, p. 57). A popular alternative is visitor self-reporting, but as explained in Chapter 1, even honest self-reports are inherently biased by self-beliefs.

Multiple-choice questions

Multiple-choice questions have been extensively used to assess content knowledge in formal settings (Allum et al., 2008; Ceuppens et al., 2018; Falk & Storksdieck, 2005; Fencl, 2010; Gormally, Brickman, & Lutz, 2012; Hill & Sharma, 2015; Hill et al., 2015; Hill et al., 2014; Kahan et al., 2012; National Academies of Sciences Engineering and Medicine, 2018).

If the questions are appropriately constructed, a multiple-choice test can be an efficient and objective method of examination that is both capable of discrimination and easy to grade (Brady, 2005). But it also needs to be carefully designed to assess knowledge without making visitors feel uncomfortable, or validity would come into question (National Research Council, 2009).

Memory and learning are functionally two sides of the same coin (Baddeley, 1997; Falk & Dierking, 2016). This implies that a multiple-choice test assessing recollection of facts can effectively assess knowledge learning.

What to ask is important to decide. There is no such thing as general public or basic knowledge (M. Burns & Medvecky, 2016); memory is always selective (Baddeley, 1997). Despite there not being a universally applicable option of what to ask, there are some science topics that it is better to avoid. For example, climate change is susceptible to polarization depending on a visitor's worldviews, rather than on facts (Kahan et al., 2012). Questions were selected such that the risk of conflict between the questions and visitor worldviews were minimal. As stated by Roos (2014), physical facts, like the relative size of electrons and atoms, should in theory be immune to religious and political motivations.

Table 3-7 shows the questions and choices used to assess scientific knowledge at Discovery World[54] (SL1 and SL2, plus SL3 as a control question) and Tūhura (SL1 to SL5,

[54] Discovery World instrument had a fourth question, 'I think electricity and magnetism are... [Closely related / Somewhat related / Totally independent / I don't know]'. It was not detected during the pilot (n=13), but after the final data collection, it was evident that the question was confusing, especially for younger children. It is best to remove inadequate items from the analysis (Frechtling et al., 2010).

plus SL6 as a control question). All the topics asked in the questions, except the control question's, were related to the content present at Discovery World's Light Zone and Tūhura's Unseen Forces Zone, as appropriate. All the items had one option that was right, two wrong, and an extra 'I don't know' (not displayed in the table). 'Don't know' options have been included previously in assessments of knowledge and counted as incorrect (Salmi et al., 2015). However, it is important to acknowledge that IDK may have a different meaning in a self-report (most likely indecision) than in a question about knowledge (either indecision or actual ignorance).

Table 3-7
Questions and answers to assess scientific knowledge

		Question	Options
Tūhura	Discovery World	SL1. Can you see electromagnetic waves?	I can see some of them
			Yes, all of them
			No, none of them
		SL2. Can atoms give off light?	Yes, they can
			No, they can't
			Yes, but only in labs
		SL3 [a]. What can be used to split white light into different colours?	Prism
			Mirror
			Camera
		SL4. What colours can you combine to form white?	Red + Green + Blue
			Cyan + Yellow + Magenta
			Red + Blue + Yellow
		SL5. Does your body give off infrared radiation?	Yes, all the time
			No, never
			Yes, but only sometimes
		SL6 [b]. Which of the following things is necessary for an aurora to happen?	An atmosphere
			Must be wintertime
			No moon in the sky

'I don't know' was an optional answer for each question.
[a] Control question in Discovery World, but not in Tūhura.
[b] Control question in Tūhura. One of the planetarium shows had a short mention to auroras, but only 19% of the visitors went to the planetarium, there were 5 shows on display and the particular show was not of the most popular. It is expected that less than 4% of the visitors had access to that information.
Questions and their correct answers are greyed out.

Scientific knowledge items were dichotomized (1 for each right answer, 0 for incorrect answers). The Kuder-Richardson Formula 20 (KR-20) coefficient is the equivalent of Cronbach's alpha for dichotomous values (e.g., right/wrong). The interpretation is also similar (Heale & Twycross, 2015; Hernández et al., 2014; Kuder & Richardson, 1937) . However, it is important to consider that although all items were related to light and electromagnetism, they cover multiple subtopics and it can't be expected that someone who knows about one, knows about all. In other words, scientific knowledge is not necessarily mathematically unidimensional, nor a concrete construct, which implies that the KR-20 coefficient doesn't need to be as high as Cronbach's alpha. For example, McGahee and Ball (2009) used this coefficient to analyse nursing knowledge, but conclude that "a reliability coefficient of greater than 0.50 can be considered a good coefficient for a nursing examination because most nursing examinations cover multiple concepts and topics" (p. 169). The KR-20 coefficient for Discovery World was .462 before the visit and .593 after the visit. For Tūhura, it was .506 before and .542 after the visit. The scientific knowledge score was obtained from summing up all correct answers (1 point each).

Modes of Learning Inventory (selected items)

The Modes of Learning Inventory (MOLI) is a 10-item five-point Likert-type scale developed by Environmetrics Pty Ltd (J. Griffin, Kelly, Savage, & Hatherly, 2005) to provide a measure of whether visitors themselves believe they have been learning and how they have been learning, rather than what they learnt (J. Griffin et al., 2005). Other authors have used MOLI (J. Griffin et al., 2005).

MOLI was designed to be conducted only once, in the post-survey. Still, out of concern for the survey length, it was reduced. Reversed items and those more complicated for children were dropped. A subset of six items was chosen (Table 3-8). As expected, it was still unidimensional (a single factor[55] explains 50% of variance) and had an acceptable internal consistency (α=.784)From the 198 visitors that filled out the instrument, 30% were

[55] Bartlett's test of sphericity $\chi^2(198,15)=319$, $p<.001$, KMO=.816.

Children, 19% Adolescents, 21% Young Adults, and 39% Mature Adults. Females represented 59%.

Table 3-8
Selected items from the Modes of Learning Inventory (MOLI)

MOLI1. I discovered things that I didn't know.
MOLI2. I learnt more about things I already knew.
MOLI3. I remembered things I hadn't thought of for a while.
MOLI4. I shared some of my knowledge with other people.
MOLI5. I found the exhibition educational.
MOLI6. I got curious about finding out more about some things.

From J. Griffin et al. (2005).

Direct self-report

Since different forms of intelligence and learning styles result in different learning outputs (Hooper-Greenhill, 2002), cognitive changes are highly individual and difficult to assess in a standardized way (National Research Council, 2009). Consequently, outcomes need to be assessed in a variety of ways (Falk & Storksdieck, 2005; National Research Council, 2009).

Directly asking a visitor if they learnt something new can be a way to quickly assess changed knowledge and understanding (Longnecker et al., 2014). That was done with the closed question 'Do you consider you learnt something at Tūhura's exhibits that you did not know before? (including any previous visits) [Yes / No / I haven't interacted with Tūhura's exhibits]'. Those that answered positively were asked, 'Can you give an example of something you learnt?' (and had a space to answer).

The rationale is that individuals are capable of understanding and self-reporting their own learning (Colliver & Fleer, 2016; Falk & Needham, 2013; National Research Council, 2009) and self-reported academic performance is sufficiently adequate for research purposes (Logi Kristjánsson, Dóra Sigfúsdóttir, & Allegrante, 2010).

It is worth explaining that this rationale is not in conflict with the previous justification for assessing scientific knowledge objectively. The need for proving there is knowledge gain objectively does not discredit the supposition of self-reporting validity. To the contrary, a positive gain objectively measured can strengthen the self-reporting assumption.

From the 369 visitors that answered the initial question, 26% were Children, 21% Adolescents, 32% Young Adults, and 20% Adults. Females represented 60%. Out of the 369 respondents, 75% said they learnt something new, 21% reported not learning anything new, and 4% didn't interact with the exhibits.

3.3.15 Survey qualitative data

Given that respondents had total freedom in their answers to open questions, it was decided to keep all qualitative data methods together in the same section, instead of trying to fit them into a single area of scientific literacy. The first qualitative method was a series of open questions that were asked throughout the four surveys[56]:

- "It was cool to learn about…" [a,b]
- "It would be cool if Discovery World / Tūhura had an exhibit about…" [a,b]
- "Do you miss anything from Discovery World?" [b]
- "Have your say! Let us know what you liked or didn't like of Tūhura's exhibits or how they can be improved."[57] [b]
- "Can you give an example of something you learnt?" [b] (for those that said they learnt something).
- "If you have any suggestions or comments, this is your space." [a,b]

Answers to open question were analysed in the search for relevant themes. Due to the similarities in the answers, some of the questions were coded under the same set of codes. More about the coding process can be found in Section 3.5.2.An interesting variation of open question that produces qualitative and quantitative data involves asking visitors what three words they would use to describe the science centre (Longnecker et al., 2014). In this research it was asked before the visit (words that describe what they expected the science centre to be) and after the visit (words that describe what they actually found the science centre to be).

Another source of qualitative data is not a method as such. Sometimes visitors—not only respondents, but people unaware of the research—approached the author to chat about the science centre. This was invaluable, because it provided spontaneous comments. It wasn't considered in the original design, but when it started to happen, it was decided to include this data.

> Sometimes order and linearity in design fail, and scholars may encounter various unpredictable events, urgent decisions, and unexpected interactions. These unanticipated hurdles can create possibilities for methodological adaptation and

[56] Questions with [a] were asked at Discovery World, with [b] were asked at Tūhura.

[57] Due to similarities in the responses, this one was coded together with 'It was cool to learn about…'

> alternative representations of research process beyond linearity and certainty. (Koro-Ljungberg, 2015, p. 82)

A consent form was prepared for these interactions, but due to the nature of the comments, it was not always possible to ask for a signature (e.g., It would have been discourteous to interrupt visitors while they were talking but their family had left them, and they needed to catch up). Mainly as a matter of courtesy, age and gender were not asked. Demographics were assessed visually for non-respondents[58] and obtained from the survey for respondents. Even though these interactions didn't involve a continuous engagement with visitors inside Tūhura, these notes still have some resemblance to ethnographic study, as they provide "unstructured data from a range of sources, for example fieldwork notes" (J. Jones & Smith, 2017, p. 98).

Lastly, the author took notes not only about visitor comments, but anything related to the data collection process (e.g. visitors expressing surprise that there was a planetarium). These first-hand data-based observations can be used as a method for data triangulation (Frey, 2018).

3.3.16 Questions and answers (Q&A) panel

Questions and Comments left at Tūhura's Q&A panel[59] were collected occasionally to find signs of engagement and learning. To avoid putting people off, the museum does not ask visitors for demographics. Sometimes visitors leave demographics information, but do so voluntarily. Based on what was written and the orthography, most of the comments seem to have been provided by young children.

3.3.17 Other methods considered

Other methods were devised or even piloted, but not included in the final data collection methods for a number of reasons, including lack of time for more data collection.

[58] Unless they themselves decide to share their age without being asked to.

[59] The Search Centre was a free-access adult-like space beside Discovery World. It featured a Q&A panel. The Search Centre was removed to enlarge Tūhura, and a new panel was placed with a new perspective. It expects visitors to pose questions that everybody else has a chance to comment on, sparking conversations, a technique recommended by Serrell (2015) and that is a great opportunity for indirect socialization (N. Simon, 2010). A visitor would not necessarily come back to read just one answer to their question, but to read other visitors' answers, leave additional questions, or answer others' questions. Overall, the panel was expected to produce a free flow of thoughts. Unfortunately, the idea was not well communicated to the visitors, and it is currently working similar to a traditional Q&A panel.

Some of them are viable for being conducted under different circumstances. They are presented and discussed in Appendix A.

It is important to mention the context of how most of these methods were trialled, as it has repercussions in later discussions about how learning occurs. The science centre has a program of sleepovers. Children come with their school group at around 4pm one afternoon and leave the next day at around 9am. They have the science centre all to themselves; the museum is closed to the public from 5pm to 10am. The methods piloted with sleepover children were conducted after their arrival (but before visiting the science centre) and before they left. In all cases but one, there was no evidence of learning at the sleepovers. Most children were between seven and nine years old (Years 3 to 5). The only group that showed a significant difference from pre to post was the one comprised of adolescents (Year 13, around 17 years old). This group created a mind map around the concept of 'Light'. Unfortunately, as it is uncommon for adolescents attend these sleepovers, this group was the smallest (N=7). Three of the children groups also created mind maps under the same methodology as the adolescents, but none showed any significant difference from pre to post.

Unlike Personal Meaning Maps, where the post-visit map uses the pre-visit map to keep updating it (Adams, Falk, & Dierking, 2003; Falk & Dierking, 2016), the pre-mind map was not returned to the groups; they started from scratch both days. The full results and analysis of these mind maps can be found in Appendix A.

3.3.18 Reliability, validity and trustworthiness

An instrument's reliability, validity, and trustworthiness, as well as the data it provides, are vital to produce robust results and to strengthen confidence in drawn conclusions. Reliability (Table 3-9) means that the instrument produces consistent results (i.e., the results can be reproduced) (Hernández et al., 2014; Neuendorf, 2016).

Table 3-9
Measures used to ensure reliability of instruments

Care taken	References
Instruments were piloted and corrections were made when deemed necessary.	(Fowler, 2013; Rennie & Williams, 2002)
Questions that required a good deal of thought or those believed to be sensitive were sent to the middle section of the surveys or interviews.	(Fowler, 2013)
Coding manuals were created, validated and consulted throughout the coding process for verification.	(Creswell, 2009; Gibbs, 2018)
Inter-coder reliability was high.	(Fowler, 2013)
Usability by children was considered in survey design. Surveys were short, uncluttered and clearly formatted.	(Adigüzel & Wedel, 2008; Borgers, De Leeuw, & Hox, 2000; Fowler, 2013)
Pre-tests and post-tests were matched one on one, allowing true comparison.	(Creswell, 2009; A. J. Friedman, 2008; Hernández et al., 2014)
Transcriptions were doubled-checked by a native English speaker.	(Creswell, 2009; Gibbs, 2018)
Missing data were managed under a set of rules and with a reliable imputation method.	(García, Luengo, & Herrera, 2015)
Items from a single scale/instrument were randomized each time (including from pre to post).	(Fowler, 2013)
Unreliable quantitative data were removed prior analysis.	See Section 3.4.2
Unreliable qualitative data were removed prior analysis.	See Section 3.5.1
Cronbach's alpha was calculated for all scales	(Field, 2013; Hernández et al., 2014)
Cronbach's alpha was calculated only after proving unidimensionality.	(Kind et al., 2007)
Cronbach's alpha was calculated before and after missing values imputation (the maximum difference was 1%).	(Creswell, 2009; Streiner, 2003; Tavakol & Dennick, 2011)
Kuder-Richardson Formula 20 (KR20) was calculated for the scientific knowledge instrument (the minimum was .462)	(Heale & Twycross, 2015; Hernández et al., 2014)
Factor analysis was conducted on scales to confirm unidimensionality (the minimum variance explained by a single factor in a scale was 50%).	(Moreiera, de Carvalho, & Horváth, 2019)
Factor analysis was conducted only when Kaiser-Meyer Olkin[60] (KMO) was at least .5	(Hair, Black, Babin, & Anderson, 2014; Kaiser & Rice, 1974; B. Williams, Onsman, & Brown, 2010)
Factor analysis was conducted only when Bartlett's test of sphericity[61] was significant.	(B. Williams et al., 2010)

Validity (Table 3-10) means that the instrument measures what it intends to measure (Hernández et al., 2014; Neuendorf, 2016).

Table 3-10
Measures used to ensure validity of instruments

Care taken	References
Internal validity was increased by being the same researcher who administered all the instruments in consistent behaviour.	(Hernández et al., 2014)
Construct validity and content validation of instruments were done by a panel of experts that include a PhD in Physics, a PhD in Science Communication, the author (candidate to PhD in Science Communication and with a background in Physics), a Science Communication Research Assistant and a group of Science Communication PhD candidates and Master's students.	(Hernández et al., 2014; Kind et al., 2007; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003)
Care was taken in the semi-structured interviews to avoid leading questions.	(Švec et al., 2017)
The right number of categories for the population's discrimination capacity was considered when creating the questions.	(Fowler, 2013)
For surveys collected at Tūhura, constructs were assessed in scale.	(Fowler, 2013)
A summary of coded results and selected quotes were sent to interviewees for member checking.	(Frey, 2018)
Each of the scientific literacy components was triangulated with different methods.	(Creswell, 2009)
The importance of verbatim quoting was acknowledged. When feasible, comments were written down by the visitors themselves (or dictated, when writing was difficult for them). When not possible, the author wrote them down as soon as possible. If in doubt of the exact words, only the idea or story was described.	(Corden & Sainsbury, 2006; Guest, MacQueen, & Namey, 2012)

Trustworthiness (Table 3-11) refers to showing the practices are credible and auditable (Rolfe, 2006). The concepts of reliability, validity and trustworthiness are interrelated (Neuendorf, 2016). Actually, "validity is dependent upon reliability" (H. H. Shettel & P. C. Reilly, 1966, p. 481). Each step taken to ensure them is listed under the heading where it has the largest impact.

---

[60] This is a measure of sampling adequacy that relates to the variation of components necessary to identify factors.
[61] Bartlett's test is similar to Levene's test, only more sensitive to departures from normality.

Table 3-11
Measures used to ensure trustworthiness of instruments

Care taken	References
Threats to validity, such as pre-testing, were acknowledged.	(Creswell, 2009)
It is acknowledged that while similarly motivated visitors tend to have a qualitatively similar experience, each visit is unique.	(Falk & Dierking, 2016)
Any unexpected results were analyzed and reported to dictate the strategy to follow.	(Creswell & Clark, 2017)
The Human Ethics Committee of the University of Otago approved the instruments and methodology used	Approvals 17/062 and D17/186 (Appendix B)
Māori Research Advisor of the University of Otago was consulted about the instruments and methodology used.	Approval 5697_19577 (Appendix B)
Participants were informed about the project and consented to participate.	(Fowler, 2013)

A note on Cronbach's Alpha

The use of Cronbach's alpha is well known for calculating a scale's reliability (a.k.a internal consistency). What is not completely understood is the interpretation of the value. One rule of thumb considers alpha above .9 to be excellent, above .8 good, above .7 acceptable, above .6 questionable, and lower values unacceptable (George & Mallory, 2009). However, the interpretation of these figures can move upwards or downwards, depending on the requirements of the field (George & Mallory, 2009). For example, increasing alpha beyond .8 decreases error very little, but if important decisions need to be made, then a reliability of at least .90 is desirable (Nunnally & Bernstein, 1994).In general, the closer Cronbach's alpha coefficient is to 1.0 the greater the internal consistency of the items in the scale (Gliem & Gliem, 2003). However, an alpha 'too high' may also suggest redundancy (Tavakol & Dennick, 2011), and some even recommend it should not surpass .9 (Streiner, 2003).

The concept of reliability in Cronbach's alpha assumes the scale is unidimensional (Cortina, 1993; Tavakol & Dennick, 2011). A measure is said to be unidimensional if the items in an instrument are measuring the same construct (Tavakol & Dennick, 2011). Although it would be very uncommon a high alpha in a multidimensional scale, a high value of alpha does not mean the scale is unidimensional (Gliem & Gliem, 2003; Kind et al., 2007). The reason being that the calculation of alpha depends on the number of items, and if they increase, so does alpha, regardless of the number of dimensions (Panayides, 2013). "Given any combination of a 2-dimensional test (whether the two factors are highly, moderately or weakly correlated) with a sufficient number of items alpha can exceed 0.70 or 0.80 or even

0.90” (Panayides, 2013, p. 694). An independent technique, such as factor analysis, can assess this (Field, 2013; Gliem & Gliem, 2003; Kind et al., 2007; Tavakol & Dennick, 2011). Factor analyses were performed to mathematically prove unidimensionality in the scales of this research. However, it’s important to clarify that in cases where the number of items is not excessive and were designed to be unidimensional, it is extremely unlikely they will have more than one factor and still a high alpha.

A note on KR20

The Kuder-Richardson Formula 20 (KR20) is a special case of Cronbach’s alpha for dichotomous data, such as the scientific knowledge instrument, where the responses can be dichotomized in right (1) or wrong/IDK (0). The interpretation of KR20 is similar to alpha.

3.4 Quantitative Data Analysis

3.4.1 Quantitative pre-processing

Ideally, data should be correct (truthful and faithful values and descriptions), unambiguous (one clear meaning), consistent (meaning doesn’t change) and complete (no missing values) (Kimball & Caserta, 2004). But real world data are often inaccurate and need to be cleaned (pre-processed) because of issues such as[62]: data entry errors [a], misspelling [b], missing values [c], inconsistent or contradictory data [d] and other anomalies, as discussed by Baskar, Arockiam, and Charles (2013) [a,c,d], Dou, Wang, and Liu (2015) [c,d], García et al. (2015) [a,c,d], Kimball and Caserta (2004) [a,b,c,d], Kotsiantis, Kanellopoulos, and Pintelas (2006) [b,c,d], Maletic and Marcus (2000) [a,c,d], Moreiera et al. (2019) [c,d], and Sidi et al. (2012) [a,b,d].

Data mining (a.k.a. knowledge discovery from a database) is the process of extracting potentially useful unknown information and patterns from data (García et al., 2015; J. Han, Kamber, & Pei, 2011). Although data mining is typically done with big datasets, several techniques were adapted to clean the survey data of this thesis:

1. Drop-out responses[63] were analysed to detect if there was a pattern that affected reliability. Since no pattern was discovered, they were removed.
2. Responses from children under 8 years old were removed[64] because the ethics approval was restricted to participants 8 and over (see Appendix B).

[62] The superscript letters connect the types of errors and their respective references.
[63] Most were from respondents that didn’t come back for the second part. The rate was 13% (35 of 262) in Discovery World and 7% (87 of 1237) in Tūhura.
[64] 14 younger children filled out the survey because their parents insisted they participate. Anecdotally, it was seen that these children did give their own answers, only needing help from parents to read faster and understand

3. Some visitors agreed to participate, but skipped sections or answered randomly. This poses a threat to validity (J. W. Osborne & Blanchard, 2011). Data quality can be improved by removing survey responses that exceed an acceptable number of missing attributes (Baskar et al., 2013; Kimball & Caserta, 2004). A method to detect these invalid responses was devised. It consists of assigning points to situations that indicated validity issues (e.g., taking too short time to fill out the survey or skipping questions) and then removing the whole response if total points surpassed a certain threshold in either pre- or post-surveys.
4. Responses that, to the researcher's best judgement, were clearly unreliable, but were not detected in the previous step were manually removed (e.g. the automatic system can't detect that 'Okay, Okay, Okay' is not a reliable answer to the question asking for three words to describe the science centre). When this happened, the entire survey response was removed from the dataset.

After cleaning, the final number of valid pre-post responses was 224 for Discovery World (Survey DW1) and 1099 for Tūhura (464 for Survey T1, 386 for Survey T2 and 249 for Survey T3).

A second pre-processing step after cleaning is data conformation[65], i.e., standardizing the data in the same format (Kimball & Caserta, 2004). The main conformation procedures applied to this research were as follows.

1. Sometimes the first and second part of the survey were completed on different iPads (e.g., visitors mistaking their iPad when coming back from the visit). They were matched later using notes taken *in situ*.
2. Time required to fill out the surveys was automatically recorded by iPads, but sometimes the equipment suffered from memory overflow and produced random figures (typically negative or ridiculously big). Those were removed.
3. Cases where visitors input data in the incorrect format (e.g., Age: "35 but look 25", instead of "35") or selected "Other" when the option they wanted was available (e.g. Residency "Other: Queenstown", instead of "Rest of Otago"), were fixed.

The last pre-processing step involved the handling of missing values. Missing values not only come from blanks, but also from I Don't Know responses (Kimball & Caserta, 2004). Inappropriately handling them may introduce bias (H. Wang & Wang, 2010).

---

some questions. But their data were not included because of limits on what the university ethics approval process permitted for this research project.

[65] Data pre-processing, cleaning, and conformation are sometimes described the same way or sharing elements. This ambiguity in terminology is not further discussed, only acknowledged in the literature. For the purpose of this research, conformation is a complementary step to cleaning, both being part of pre-processing.

A simple and common approach is discarding all instances with at least one missing value, but it may produce loses of important data and bias the results (García et al., 2015; Moreiera et al., 2019; Roos, 2014; Rubin, 1976). Replacing the missing values by fixed values is not necessarily a better option (Moreiera et al., 2019). Instead, a reasonable estimate can work better than the previous alternatives (García et al., 2015). Expectation Maximization (EM) was chosen as the method to cope with missing values, as it estimates the parameters of an incomplete dataset by maximizing the likelihood of fitting a probability distribution iteratively[66] (García et al., 2015).

Imputation requires missing data to be either missing at random (MAR) or missing completely at random (MCAR). MAR means the missing values depend on the observed values, but not on the missing values. Missing completely at random (MCAR) is a special case where missing values do not depend on either the observed values or the missing values (García et al., 2015; Tsikriktsis, 2005).

All scales were tested for MCAR using Little's MCAR test (Little, 1988). When a set of missing values fails to be MCAR, there is no statistical test to know if it is still MAR, but it can be directly analysed searching for patterns (bias) (H. Wang & Wang, 2010). Item range variation, standard deviations, percentage of missing values and plotting of missing values were the methods used to holistically decide if non-MCAR data were still MAR. Missing data in six of the scales were MCAR and five were not. Since no evidence of patterns was found in the latter, they were assumed to be MAR.

Imputation was restricted to responses with up to 40% missing values (blanks + IDK) in each given scale (pre and post considered separately[67]). If more were found, the complete scale (pre and post) was deleted for that participant. Expectation Maximization[68] helped to input 76 of 170 blanks and 315 of 590 IDK. Eighty-five scales were deleted for having too many missing values.

The 3-item ranking of Mathematics, Explanation, and Experiments is not a scale. Still, a type of imputation was done. If one blank was present (either before or after the visit), the only remaining possible value was input. If there were more missing values, then the whole response (pre and post) was deleted. Twenty-six imputations were done in the ranking instruments and 24 complete responses (with a total of 84 blanks) were deleted.

[66] The most accurate technique is Multiple Imputation, which is similar to Expectation Maximization, but it produces a series of probable sets. It was not chosen because each set needs to be analysed independently, making it unfeasible when there are many scales.
[67] That is, up to two missing values out of five, six or seven values, depending on the scale.
[68] Mersenne Twister random numbers generator (Matsumoto & Nishimura, 1998) was used.

No imputation was done in the multiple-choice instrument to assess scientific knowledge, but eight of 464 responses were removed for having too many skipped items (notice that for this purpose, IDK was not considered missing). Pre-processing steps can be found in full in Appendix D.

3.4.2 Descriptive analysis

Descriptive statistics refers to ordering and summarizing data in tables and charts to facilitate their visualization and interpretation (Cumming, 2013; Diggle & Chetwynd, 2011; Marshall & Jonker, 2010; Moreiera et al., 2019). It also includes numerical figures, called descriptives, that summarize information about the data.

When a distribution is normal, mean and standard deviation are sufficient descriptives (Rendón-Macías, Villasís-Keeve, & Miranda-Novales, 2016). If the distribution is not normal, it is more illustrative to present medians and inter-quartile ranges (Hernández et al., 2014; Rendón-Macías et al., 2016). Plots, charts and diagrams presented in this thesis were created in Excel™ 365, except those described below.

LOESS fit

When we fit a straight line, a quadratic curve, or an exponential to a data set, we are assuming beforehand the data behave the way we are approximating them. To avoid trying all possible options, we can plot points and visually decide what's the best fit. But what to do when the tendency is not obvious to the eye?

A LOESS fit (Locally Estimated Scatterplot Smoothing, a.k.a. LOWESS, Locally Reweighted Scatterplot Smoothing) is similar in nature to a linear regression, but instead of producing a single and linear regression from all data points, it creates multiple weighted local linear regressions around each point by using a subset of n neighbouring points (Cleveland, 1979; Cleveland & Devlin, 1988). In other words, consider a single point (x_i,y_i) from the data set and let's call it reference point. The LOESS fit takes only the nearest neighbouring points $(x_{i\pm n},y_{i\pm n})$ and uses them to approximate a linear regression. The closer a neighbouring point is to the reference point, the more weight it is assigned to it. This linear regression has an associated linear equation f_i. This process is repeated to all of the points. The new reweighted LOESS data set is composed of the points $(x_i,f(x_i))$. By joining this points we obtained a smoother fitter curve. The smaller the neighbourhood, the more sensitive the curve is. If the neighbourhood includes all the data set, the result of LOESS is just a usual linear regression. If the neighbourhood is too small, an over-fitting issue may arise due to noise. The larger the n, the less inclusion of noise, but there is a risk of lack of fit[69]. LOESS fits were calculated using SPSS™ v25.0 and the size of the neighbourhoods were found empirically.

To avoid determining a neighbourhood that varies depending on the size of the data set, the smoothing parameter, α, determines the percentage of points in the neighbourhood with respect to the total number of points, typically falling between 0.40 and 0.80 (W. G. Jacoby, 2000). In this case, α=.70 was selected as it produced the clearest results. Epanechnikov was chosen as the kernel (Epanechnikov, 1969).

Although the LOESS fit is merely descriptive and does not produce a correlation coefficient as the linear regression would, it is extremely useful to detect how data behave. LOESS doesn’t need the user to decide beforehand what mathematical relation will fit the data, it comes naturally. Moreover, if more than one relationship is present in the data, the traditional approach of one-fit won’t be useful, but the LOESS will able to detect relationships by zones, as it will become clearer in sections 4.2.3 and 5.3.3.

Structural Equation Modelling (SEM)

SEM is a multivariate statistical analysis technique used to analyse structural relationships, such as the relationship among more than two constructs. SPSS AMOS™ v25 was used to produce both a plot and set of correlations between each pair of constructs. Since it is presented visually in this thesis, it is kept in this section.

Learning flow diagram

This diagram was created by the author to visualize how scientific knowledge learning happens. Since it is a new visualisation, it had to be drawn manually in Paint.Net™ v4.1.6. Figure 3-4 exemplifies the concept. The diagram can be done in several ways; here, all the scientific knowledge items were pooled in a single set. The way to read the learning flow diagram is as follows: circle diameters are proportional to the number of answers that didn’t change from the pre to the post. For instance, in Figure 3-4, 21% of pooled answers from respondents were initially correct before the visit and remained correct afterwards. Summing up the three circles gives a total of 63%, which is the percentage of stable answers.

[69] If n=N, then LOESS would be equivalent to a simple linear regression.

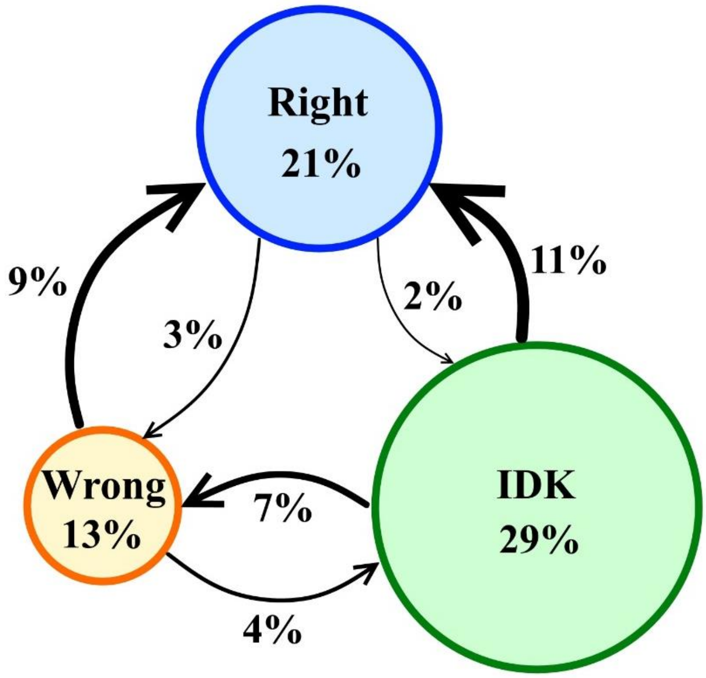


Figure 3-4. Example of a learning flow diagram. It shows how answers about scientific knowledge changed from before to after visiting a science centre.

But learning doesn't come with stability; learning comes with the change of knowledge and understanding when the visitor's mental framework is challenged (Feher, 1990; Krishnamurthi & Rennie, 2012). The arrows show how answers moved among the three options. The width of each arrow is proportional to the number of answers that changed from pre to post. The direction of the arrow explains from what and to what group they moved. For instance, in Figure 3-4, 11% of the responses were IDK before the visit and changed to the right answer after the visit.

The pre and post results show there were 26% right answers before (21% from the circle, plus 2% and 3% of the outgoing arrows) and 41% after (21% from the circle, plus 9% and 11% of the incoming arrows). A learning flow diagram draws a larger picture. It not only shows the outcome of learning, but how learning happened.

The descriptive statistics used to describe the results in this thesis depended on the nature of the data. For parametric data, Mean (M), Standard Deviation (SD) and Confidence Interval (CI)[70] are reported. The corresponding descriptives for nonparametric data are Median (Mdn), Interquartile Range (IQR) and Confidence Interval (CI)[71].

[70] When plotting, error bars will correspond to confidence intervals.

[71] The equivalent of confidence intervals (M-SD, M+SD) is (1st quartile, 3rd quartile). It does not always coincide with the range given by Mdn ± IQR. It will be acknowledged whenever that happens.

3.4.3 Parametric and non-parametric inferential statistics

Inferential statistics use data from the sample group to make generalizations that can be applied to a larger population (Cumming, 2013; Marshall & Jonker, 2010). Hypotheses can be tested parametrically or nonparametrically. Parametric tests are more powerful than non-parametric tests (Field, 2013). But unlike nonparametric tests, parametric tests require observations to be independent, the model to be linear, and the distribution of means to be normal (García et al., 2015). A few tests also can't handle data unless the variance is approximately homogeneous (homoscedasticity) (García et al., 2015). Independence and linearity are usually taken for granted. When a test can´t handle heteroscedasticity, a test can be conducted to calculate it. Therefore, normality is the most critical of the conditions.

A common misconception is that the sample itself needs to be normally distributed, but it is the sampling distribution of means that needs to be normal (Field, 2013). Fortunately, the central-limit theorem states that the distribution of means for an infinite sample forms a normal distribution regardless of the shape of the sample and normality can be assumed (Field, 2013). This distribution can be approached with a "sufficiently large" sample size, though the minimal recommended value varies: 15 (Hernández et al., 2014), 20 in light-tailed distributions (Field, 2013), 40 (Diamond et al., 2016), 64 (Soper, 2019)[72] and 160 in heavy-tailed distributions (Field, 2013). The most common value is 100 (Field, 2013; Hernández et al., 2014; Wilcox, 2010).

This lower limit becomes trickier with Likert-type scales. It is true that these ordinal scales are usually associated with numbers (e.g., small=1, medium=2, large=3), but these numbers are not figures, they are just an identification code (Hernández et al., 2014; Marshall & Jonker, 2010; Moreiera et al., 2019), like the number on a bank card. On the other hand, ordinal data are qualitative in nature (Villasís-Keever & Miranda-Novales, 2016), and qualitative variables classify the relation between individuals and phenomena through attributes (Diggle & Chetwynd, 2011), not quantities.

Even so, Likert-type scales are commonly tested parametrically (Hernández et al., 2014; Moreiera et al., 2019). A number of researchers consider parametric tests to be robust enough to handle Likert-type scales without biasing the conclusions drawn from the tests (Boone & Boone, 2012; Carifio & Perla, 2008; Derrick & White, 2017; Murray, 2013; G. M.

[72] This author provides a tool with which this value was calculated.

Sullivan & Artino, 2013), even in the absence of normality (Norman, 2010; Wadgave & Khairnar, 2016).

Flow diagram to decide whether data can be treated parametrically or not

Regardless of the power of parametric tests to handle non-normal data, whether Likert-type scales can or cannot be considered interval data is debatable[73]. The concept of normality doesn't make sense with ordinal data. Appendix E presents an in-depth discussion about the topic, concluding that Likert-type data can be considered a good approximation of interval data. But since, by definition, it is not interval, heavier restrictions were assigned to Likert-type scales and other ordinal data to be analysed with parametric methods.

The author created a flow diagram to decide when to consider collected data to be parametric (Figure 3-5). In this context, parametric data is understood as that which can be analysed parametrically.

It was mentioned above that normality is a critical condition in deciding if a set can be treated parametrically (P) or not (NP). But normality is not dichotomous. To decide if data are 'normal enough' is difficult. There are tests to assess it, such as the Kolmogorov-Smirnov test[74] (Field, 2013; García et al., 2015). However, its results alone can be deceiving, producing false negatives (Field, 2013). Micceri (1989) analysed 440 distributions and all failed the test. Other options to complement the test and holistically determine if data are normal enough include Histograms and Skewness[75] (Field, 2013; García et al., 2015; Ho & Yu, 2015; Micceri, 1989) and Q-Q plots (Field, 2013; García et al., 2015). The value of skewness that can be considered acceptable depends on context (Ho & Yu, 2015), but up to 0.30 can be considered acceptable, with 0.31 to 0.70 moderate and above 0.71 extreme (Micceri, 1989). How normal data needs to be is to be assessed holistically by the researcher, and sample size is taken into consideration. For example, it is not the same if the sample size of some interval data is 99 or 20 (see Figure 3-5). The former case would need to be minimally normal to fit, while the latter would need to resemble almost a perfect Bell curve for a parametric test to be valid.

[73] The main problem is whether anchors can or cannot be considered equidistant. Notice that there are ordinal data that cannot be considered parametric in any case, for example, age ranges, because they are not equidistant.
[74] In this case, it was used with Lilliefors significance correction.
[75] Also with kurtosis, but its interpretation is more difficult (Ho & Yu, 2015; Micceri, 1989).

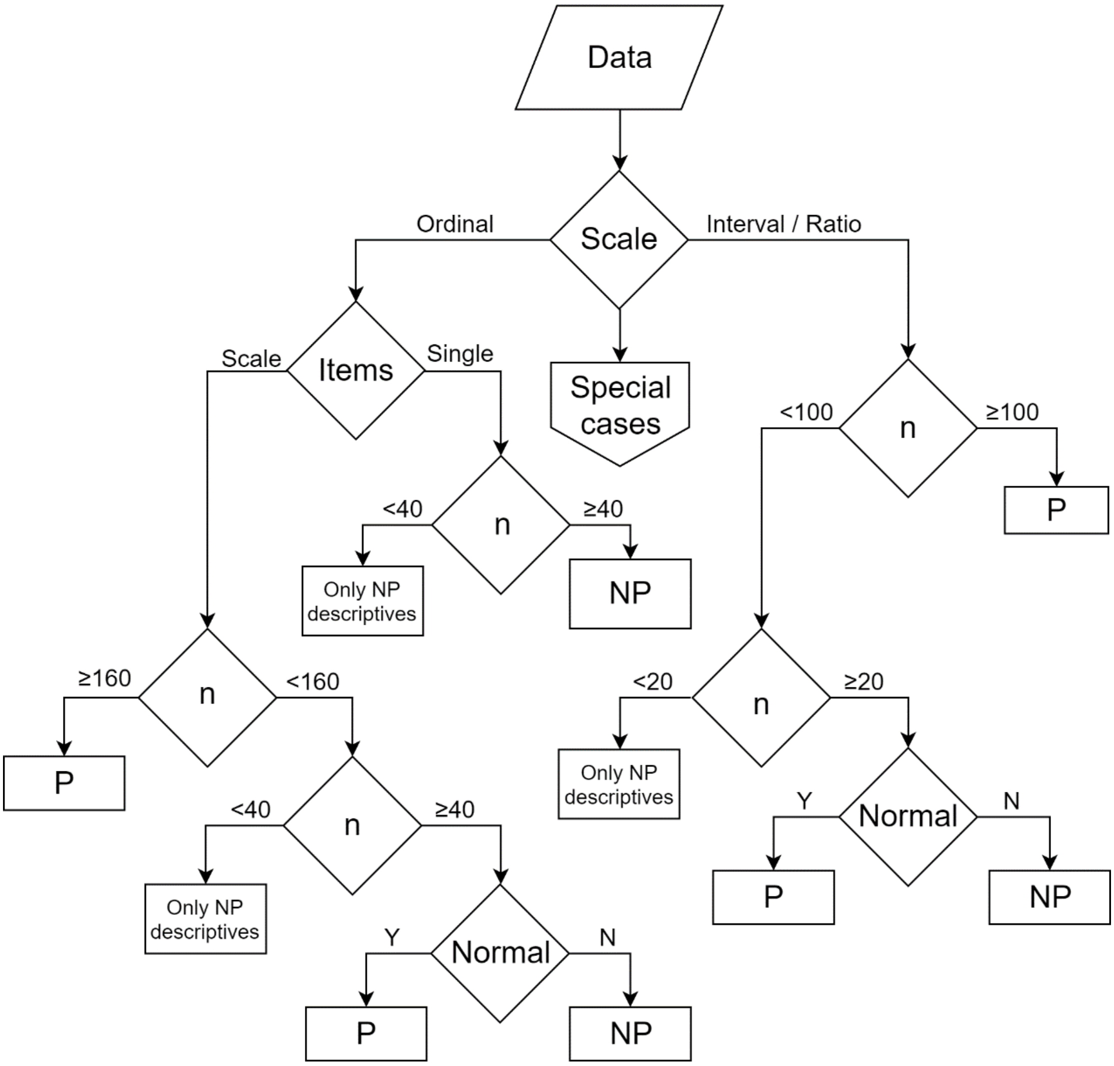


Figure 3-5. Flow diagram to decide the type of data. P stands for parametric, NP for nonparametric and n for the sample size. See the special cases in text.

The special cases signalled in Figure 3-5 can be divided in two. First, to determine correlations, the needed sample size depends on the expected effect size and the number of predictors, regardless of whether the data are interval/ratio or ordinal. For example, expecting[76] a medium effect size, the minimum sample size would be 55 for one predictor and 68 for two predictors (Field, 2013).

Second, all categorial data are automatically nonparametric because normality makes no sense in categories. Thus, a minimum sample is not defined per se. However, Chi-Square tests[77] require at least five cases in at least 80% of the cells[78] (Field, 2013; Kim, 2017;

---

[76] This is tricky when you are exploring and don't have idea how big the effect size will be. Medium sounds reasonable in those cases.

[77] Such tests are considered the best approximation for contingency tables; the most used are Chi-square goodness of fit and Chi-square test for independence.

[78] The expected value for each cell is: (row total)(column total)/overall total, not the observed number in each cell.

McHugh, 2013). This condition is usually satisfied when the sample size is at least equal to five times the number of cells (McHugh, 2013)[79]. If the minimum number of cases is not reached, Chi-Square can be replaced by Fisher's exact test (Field, 2013; Kim, 2017; Lydersen, Fagerland, & Laake, 2009; McHugh, 2013).

Flow diagram to choose a statistical test

Once data have been catalogued as either parametric or nonparametric, the decision of which test to conduct can be decided according to Figure 3-6. This diagram is also the creation of the author. To test the homogeneity of variances when it is a condition, Levene's test can be used (Field, 2013).

[79] E.g., 2x2 would require $n \geq 20$, but this is just to sample goal, the actual number of cases in each cell needs to be checked before conducting the actual test.

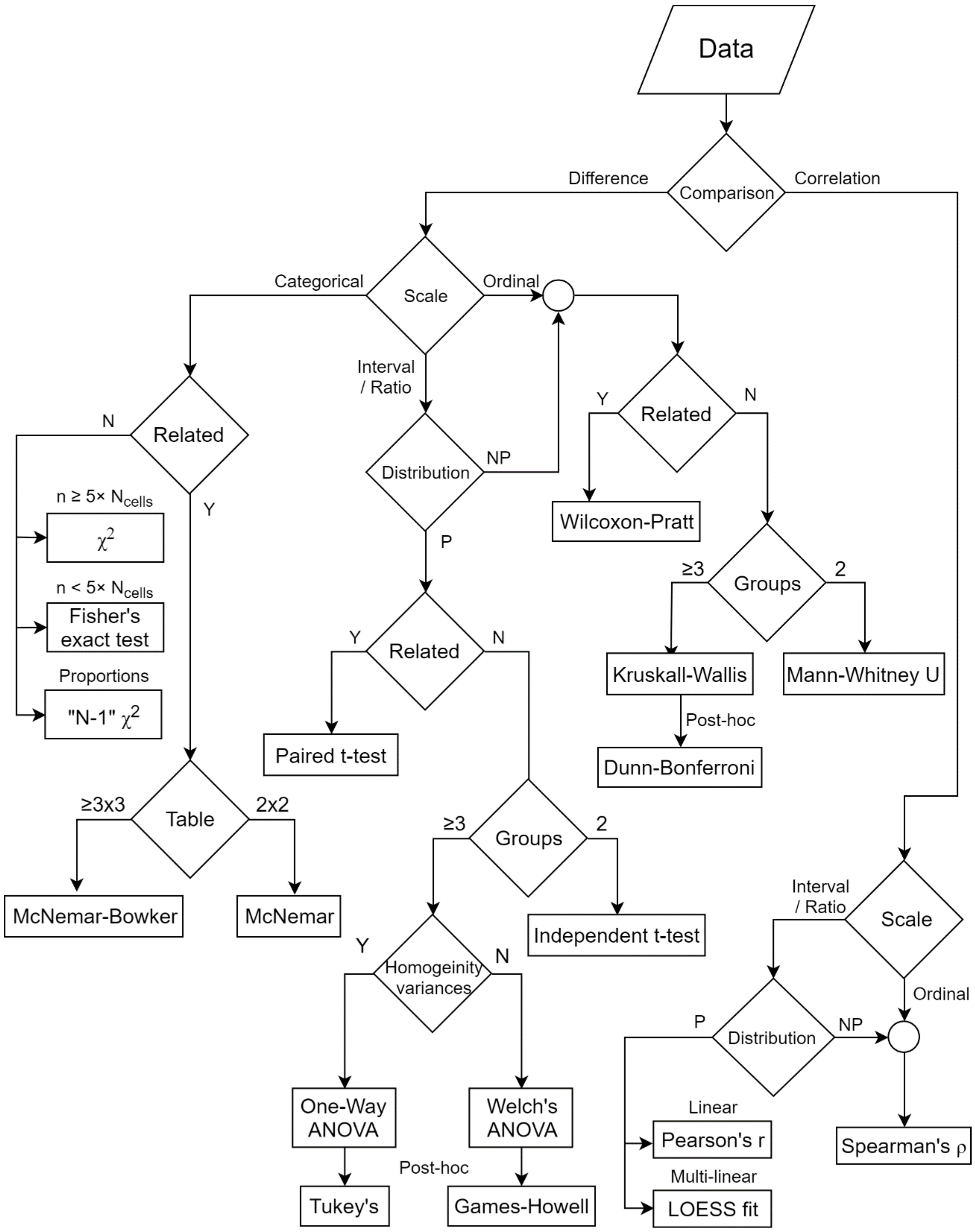


Figure 3-6. Flow diagram to choose the appropriate statistical test. P stands for parametric data distribution, NP for nonparametric, n for the sample size, Y for yes and N for no.

Software

Most tests were conducted using SPSS™ v25.0. Wilcoxon-Pratt test was done in R3.4 (GUI RStudio v1.1). 'N-1' Chi-square test (Campbell, 2007; Richardson, 2011) and proportion confidence interval (Altman, Machin, Bryant, & Gardner, 2000) were tested online with the software developed by MedCalc Software (2019). Effect sizes and confidence intervals were mainly calculated on a handheld calculator and Excel™ 365.

3.4.4 The null-hypothesis statistical testing

The null-hypothesis statistical testing (NHST) paradigm is usually employed to accept or reject a hypothesis depending on whether $p<\alpha$. The p-value (or simply p) is the probability of getting observed data, given that the null hypothesis is true. A type I error is the rejection of a true null hypothesis, i.e., a false positive, and α is the probability of making that error (Gigerenzer, 2004; Haller & Krauss, 2002). Usually, α is set to .05 or .01)[80]. However, a significance test cannot actually prove or disprove a hypothesis, only provide probabilistic information (Haller & Krauss, 2002). Type I errors may still appear with a sufficiently large sample (G. M. Sullivan & Feinn, 2012). What needs to be communicated is information, not a dichotomous decision (Cumming, 2013; Field, 2013; Gigerenzer, 2004).

An exact value of p is preferred over a general statement of passing or failing the α condition (American Psychological Association, 2010). Second, the p-value provided in the NHST can be complemented with the effect size and the confidence interval[81] (Cumming, 2013; Field, 2013; Gigerenzer, 2004; Preacher & Kelley, 2011).

3.4.5 Confidence interval

A confidence interval (CI) is an interval estimate of the range of upper and lower statistical values that are consistent with the observed data and likely contain the actual population mean (Creswell, 2009). The $95\%\ CI$ can be calculated for parametric data as $[M - 1.96 \times SE, M + 1.96 \times SE]$, where SE is the standard error and M the mean (Cumming, 2013).

Although relatively imprecise, 95% CI can also be calculated for nonparametric data. Instead of finding the interval directly, the positions of the lower and upper bounds in the ordered data are found first. These positions are found by rounding up to the next integer from the following values (n is the sample size) (Bland, 2015):

$$\left(\frac{n}{2} - \frac{1.96\sqrt{n}}{2}, \frac{n}{2} + \frac{1.96\sqrt{n}}{2}\right)$$

[80] These values are not strict thresholds coming out from a mathematical theory, but simply historical suggestions (Field, 2013).

[81] Cumming (2013) proposes ES and CI as alternatives to the NHST, i.e., to replace the p-value. However, the author considers more useful to use them as complements, rather than substitutes.

The actual confidence interval is obtained by replacing the integer values above by the ranked values in the corresponding positions.

Notice that CI is actually a range (min, max), but given than the mean and median should still be reported, CI is reported in this thesis as a symmetrical value around the mean or median, i.e., $\mathrm{Mean, Median} \pm \mathrm{CI}$. As with the parametric case, most of ranked CI values are symmetric around the median. In the few cases where they are not symmetrical, the highest distance from any of both bounds to the median is taken[82]. In the case of a single item, the CI is rounded to the next higher integer.

When dealing with proportions (percentages), the CI is calculated with the Wald method (a.k.a. asymptotic method) based on a normal approximation[83] (Brown, Cai, & DasGupta, 2001), using the online tool provided by Sergeant (2019).

3.4.6 Effect size

The effect size (ES) uses the expected difference in means between groups to identify the strength of the conclusions (Creswell, 2009). Unlike the p-value, an effect size is independent of the sample size (G. M. Sullivan & Feinn, 2012). While it is still not a generalized practice to report effect sizes (Fritz, Morris, & Richler, 2012), it is advisable to report some kind of effect size, even for non-significant results[84] (American Psychological Association, 2010; Grissom & Kim, 2005; Preacher & Kelley, 2011; Thompson, 2007).

Effect sizes associated with statistical tests

There are multiple methods for calculating effect sizes. Some tests have more than one equally valid effect size. Table 3-12 summarizes the most common association between effect sizes and tests.

Some tests don't have a traditionally associated effect size, like the McNemar test (P. Allen, Bennett, & Heritage, 2018). However, the probability that the event occurs divided by the probability that it does not occur, called Odds Ratio (OR), is commonly used as a summary measure in meta-analysis (Fagerland, Lydersen, & Laake, 2015) and can be used as an effect size for the McNemar test (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cleophas & Zwinderman, 2016; NCSS, 2019a).

[82] The non-symmetrical variations are usually very small. Only if the variation were large would the full (min, max) format be reported.
[83] There are better approximations, but this is the only symmetrical one. The difference between this and the one with the most exact calculations is minimal.
[84] As explained above, accepting or rejecting a hypothesis shouldn't be a dichotomous decision. The interpretation of p=0.051 and p=0.900 is completely different, regardless of both being catalogued as not significant in the dichotomous version of the null hypothesis testing.

Table 3-12

Effect size recommended for some common statistical tests

Effect size	Statistical test	Reference
Cohen's d	Paired samples t-test, Independent samples t-test	(Field, 2013)
r	Pearson's r, Mann-Whitney U test, Welch's ANOVA, non-parametric tests	(Field, 2013)
R^2	Coefficient of determination R^2	(Field, 2013)
r_s	Spearman's rho	(Field, 2013)
η^2	One-way ANOVA	(Cumming, 2013)
φ (w)	2x2 Chi-squared tests	(Cohen, 1988; Field, 2013)
Cramer's V	Chi-squared tests larger than 2x2	(Cohen, 1988; Field, 2013)
Odds ratio (OR)	McNemar test, "N-1" Chi-squared test	(Borenstein et al., 2009; Cleophas & Zwinderman, 2016; NCSS, 2019a)
Discordant proportion ratio sum (DPRS)	McNemar-Bowker test	(NCSS, 2019b)

Note: Many effect sizes can be used for each test (Cumming, 2013). This table shouldn't be read as a one-on-one relation, but as a guide of what effect size is most commonly reported in each case, according to the references and experience of the author.

Using measures from the thesis as an example, if visitors move between Inaction (I) and Interaction (A), the Odds Ratio (OR) is:

$$Odds\ ratio = \frac{I \rightarrow A}{A \rightarrow I}$$

The OR is actually a test, similar to chi-square test applicable to cross-tabs, but can also be interpreted as relative risks (Cleophas & Zwinderman, 2016). It can be seen as the chance of getting a result; e.g., in sports betting, what are the odds that one team beats another team. However, the fact that the OR is unbounded, and it is not easily found in the literature, might make it harder to easily interpret.

For the McNemar-Bowker test, the Discordant Proportion Ratio Sum (DPRS) was deemed more appropriate (NCSS, 2019b). It is defined as:

$$DPRS = \sum_{i<j} \frac{\left(p_{ij} - p_{ji}\right)^2}{p_{ij} + p_{ji}}$$

where p_{ij} is the proportion of column *i* row *j* (number of pairs in *i,j* divided by the total number of pairs). But that calculation is equivalent to:

$$DPRS = \frac{\chi^2}{N}$$

where $\chi 2$ is the McNemar-Bowker statistic and N is the total number of paired observations (NCSS, 2019b).

For any other nonparametric test with no effect size, the correlation coefficient (r) can be used (Fritz et al., 2012; Tomczak & Tomaczak, 2014):

$$r = \frac{Z}{\sqrt{N}}$$

where Z is the statistic and N is the sample size[85].

Strength of effect sizes

The strength of the different effect sizes is presented in Table 3-13. The values for small, medium and large are 'around'[86] the values in the table (Cohen, 1988, 1992; Durlak, 2009; Field, 2013). It is important to consider that "the terms 'small', 'medium', and 'large' are relative, not only to each other, but to the area of behavioural science or even more particularly to the specific content and research method being employed in any given investigation" (Cohen, 1988, p. 25). For example, Hattie (2009) proposes an adaptation to Cohen's d in education, where d between 0.0 and 0.15 is the zone of developmental effects (what students could probably achieve by themselves, with no school system), d between 0.15 and 0.4 the zone of teacher effects (accomplishment in a typical year of schooling), d > 0.4 would be the zone of desired effects (achievements above average) and d > 0.6 would be considered excellent.

To the best knowledge of the author, there is no direct interpretation of what value could be considered small, medium or large for DPRS and OR. However, Borenstein et al.

[85] The original formula calls the sample size 'n', but changed because it is more common that N is the sample size and n is the size of any sub-sample.
[86] There is a lot of misinformation in the literature around citing some of Cohen's suggestions incorrectly, including the interpretation of the effect size. For further discussion, see Appendix E.

(2009) give a transformation for OR that can be used in conjunction with the calculator produced by Lenhard and Lenhard (2016) to find equivalent values for OR[87].

Table 3-13
Interpretation of the strength of effect sizes

	d	r (w / φ)	r^2 (η^2)	Cramer's V	OR	DPRS
Small (S)	0.2	0.1	.01	0.07	1.4	NA
Medium (M)	0.5	0.3	.09	0.21	2.5	NA
Large (L)	0.8	0.5	.25	0.35	4.3	NA

Equivalent effect sizes are presented in parentheses.
Values come from J. Cohen (1988), except for OR and DPRS (see text).

Notice that the coefficient of determination (r^2) can be directly obtained from the correlation coefficient (r), but the interpretation is different. The correlation coefficient measures the strength of association in a linear relationship between two variables, while the coefficient of determination estimates the proportion of variance of one variable explained by the other variable (Field, 2013). For example, r^2=.350 means that 35% of the variance in the dependent variable is explained by the independent variable. In natural sciences, where lab control minimizes the effect of exogenous variables, r^2=.350 would be considered rather small (weak correlation), but humans are more complicated and the interpretation of their effect size should be read from Table 3-13. "A low value of R^2 indicates merely that the dependent variable is affected by a host of other factors in addition to the ones considered in the analysis" (Moksony, 1990, p. 3)[88]. In other words, even though r^2 indicates the proportion of variance explained by an independent variable (statistical explanation), it doesn't give a substantive explanation because it is affected by a number of factors, such as the magnitude of the effect (Moksony, 1990).

Confidence intervals of effect sizes

Lastly, although it is neglected most of the time, effect sizes themselves have confidence intervals (Fritz et al., 2012). For this research, they were calculated only for Cohen's d (Hedges & Olkin, 1985)[89]:

[87] Ferguson (2009) has recommendations for the OR, but in Appendix E it is discussed why the author does not consider it reliable.
[88] The difference between r and R is that r is used exclusively for simple linear regressions, while R is more general, including the case of multiple regression. In this case, where there is only one independent variable, r and R are equivalent.
[89] Confidence intervals for other effect sizes and the reason for not using them can be found in Appendix E.

$$95\%\ CI\Delta = d \pm \sqrt{\frac{n_a + n_b}{n_a n_b} + \frac{d^2}{2(n_a + n_b)}}$$

where n_a and n_b are the sample sizes of the groups (at least 10 per group).

This confidence interval will be reported as a single figure and represented by d_{CI}.

3.5 Qualitative Data Analysis

3.5.1 Qualitative pre-processing

Qualitative data need to be cleaned to ensure quality (Chu, Ilyas, Krishnan, & Wang, 2016). Methods used in data mining can be applied to text analysis (Dou et al., 2015; Weiss, Indurkhya, & Zhang, 2010) to normalise, clean, disambiguate and define what can be considered a word (Adda, Adda-Decker, Gauvain, & Lamel, 1997; Mikheev, 2000).

Three words to describe the science centre

Despite restricting data by asking for only three words, this data required significant pre-processing. The first recommended step is to extract words from a stream of characters in what is called tokenization (Moreiera et al., 2019). Nonstandard tokens can come from a mixture of unintentional misspellings and intentionally-created tokens, heavily influenced by how informal writing has evolved in social media (F. Liu, Weng, & Jiang, 2012). Some examples are lengthening the words to express sentiments (e.g., “Cooooooooolllll!!!!!!”) (Brody & Diakopoulos, 2011), shortening them to speed texting up (e.g., “asap”) (Crystal, 2008), or phonetic substitutions as a form of identity (e.g. “’wuz up bro”) (Hassan & Menezes, 2013).

Stemming and lemmatization are the most popular ways to reduce variations in tokens (e.g., plurals and verb inflections) (Moreiera et al., 2019). Stemming attempts to reduce a word to its root by chopping off the ends of the words. Lemmatization refers to finding the morphological base of the word by removing inflectional endings only and to return a base form that can be found in the dictionary (P. Han, Shen, Wang, & Liu, 2012; Manning, Raghavan, & Schütze, 2008). For instance, the stem of ‘was’ is ‘wa’, whereas its lemma is “be”. Both methods were tested (see Appendix D), but both produced unsatisfactory results (e.g., they were unable to recognise that ‘Funny’ can be related to ‘Fun’).

State-of-the-art automatic correction software is approaching 100% accuracy in some scenarios (Pramanik & Hussain, 2019; Zhang et al., 2019), but no algorithm gives a 100%

output (Jivani, 2011). It is still humans who can most easily interpret errors in a message (Zhang et al., in press). Since “a combination of automated and manual techniques is a potentially more useful approach” (Clark & Araki, 2011, p. 3), this was the approach used.

Stop words were removed (e.g. ‘the’, ‘a’) (Manning et al., 2008; Moreiera et al., 2019); tokens were canonicalized (e.g. U.S.A. was replaced by USA) (Manning et al., 2008); British English was set as preferred spelling (e.g. ‘colour’ instead of ‘color’) (Manning et al., 2008), phrases became hyphenated words (e.g., ‘The / Best / EVER!!’ became ‘The-best-ever’) to be discriminated (Kay & Röscheisen, 1993; Koster, 2004). Then tokens were manually stemmed, and equal stems were replaced by the lemma word (full word) with highest frequency. (Enideo, 2019) was used to count frequencies. More can be found in Appendix D.

Quotes pre-processing

Minor spelling and grammar were fixed in quotes from open questions (Metzler, 2017). Changes are signalled in brackets. For readability reasons, changes in letter case and punctuation are not shown. For example, “No i oike the new exhibits.” became “No, I [l]ike the new exhibits.” Misspellings that appeared intentional were not corrected, such as lengthening a word to express excitement. Clarifications added by the researcher are presented in brackets. For example, “Need to reinstate small SOUNDPROOF [uppercase in original]”.

Audio from interviews and focus groups was simultaneously recorded using a hand-held recorder and a smartphone in Discovery World. A number of Discovery World recordings had low quality. Audacity 2.2.2 was used to improve their quality by removing ‘clips’, reducing noise, adjusting the equalization and increasing the volume. In Tūhura, a lavalier microphone was added to the smartphone, which produced higher-quality recordings. Transcriptions were cleaned from superfluous words and sounds, such as ‘um’ and ‘you know’ (Corden & Sainsbury, 2006).

Quoting

Survey quotes include basic demographics in the format (T/DW, M/F, age), where T stands for Tūhura, DW for Discovery World, M for Male, F for Female and age is a continuous value in Tūhura, but one of the four ranges (8-12, 13-18, 19-40, 41+) in Discovery World. For example, (T, M, 35) means 35 year-old male from Tūhura. (DW, F, 8-12) indicates the quote comes from a girl between 8 and 12 years old surveyed at Discovery World.

Quotes from interviews include a code in the form of SC# or DM#, where SC stands for Science Communicator, DM for Decision Maker and # is the number of the participant. For example, DM5 means that the quote comes from Decision Maker number five. Focus group quotes include a code in the form of C/A(DW/T), where C stands for Children, A for Adolescents, T for Tūhura, and DW for Discovery World. For example, C(DW) means that the quote was made by a child in Discovery World's focus groups.

3.5.2 Coding

Due to the categorial nature of qualitative data, analysis requires retaining and describing these data with primarily inductive strategies (Frey, 2018). Coding is a popular method to quantify qualitative data. It involves creating codes and counting the number of times they occur in the text (Creswell, 2009; Hernández et al., 2014; Rossman & Rallis, 1998; Saldaña, 2015).

Coding in surveys

Open questions in surveys were mainly analyzed through descriptive coding. This coding methodology is especially useful to find topics by summarizing the topic in qualitative data with a word or a short phrase (Saldaña, 2015). After a set of categories were defined, they were summarized in a smaller set by finding themes. This process is known as pattern coding (Saldaña, 2015) and was conducted iteratively and being revised by a panel of experts (the main supervisor, a research assistant, and a group of postgraduate students). To organize the themes and categories, a coding manual (a.k.a. codebook) was prepared with explanations, examples and exclusions for the codes. The goal of a coding manual is to create a complete and unambiguous manual such that all coders code in the same way regardless of their individual differences (Neuendorf, 2016; Saldaña, 2015). Coding manuals can be found in Appendix F.1.

Coding in interviews

Analysis began with the author transcribing all recordings first. During this first stage, in vivo coding was done. In vivo is an initial coding method where worldviews are detected in transcripts by highliting verbatim mentions (Saldaña, 2015).

After the author completed the transcriptions and found codes given by verbatim mentions, an external transcriber whose native language is English double-checked the transcriptions and corrected any mistakes (Corden & Sainsbury, 2006). Having ready a final version of the transcription, a second cycle of coding was performed to help identify main topics and points of convergence. This time coding came through pattern coding, which is a way of grouping summaries into a small set of themes (Saldaña, 2015). The patterns (themes) found by the author were continuously revised and improved by the same panel of experts mentioned above until a coding manual was approved by all (Appendix F.2). The codes found during the in vivo coding was useful again once the pattern coding was completed to identify the most suitable quotes to supports the coded patterns. Extracted information and selected quotes were sent to interviewees for member checking to ensure the given interpretation was an accurate representation of their views (Creswell, 2009).

A third coding method used with interviews was magnitude coding. In this case, the intention is to supplement a code by adding a symbolic subcode related to the intensity of the code. For example, how much staff knew about the Dodd-Walls Centre was coded this way in low, medium and high (see Appendix F.2). The author was fully in charge of completing the analysis once the group of experts considered it reliable.

Coding in SWOT focus groups

Analysing data from a SWOT focus group was done a bit differently than other qualitative methods present in this research. Analysis in this case was conducted in three stages. The first one was done by each focus group participants. When they attended the science centre, they analysed the exhibits and wrote down their individual SWOT analysis. In the second stage, individual opinions were discussed in a focus group, leading to a collaborative SWOT analysis. The last stage is divided in substages. First, the author transcribed audio recordings and during this stage the researcher conducted evaluation coding. This type of coding is especially useful to detect judgements. Evaluation coding can employ an amalgam of magnitude coding (by distinguishing between positive [+] and negative [-] comments), descriptive coding (noting the topic) and in vivo coding (no note the specific comment) (Saldaña, 2015). For example: [-]Floating in Copper: “annoying to figure out how it works”, or [+]Torque Table: “so many things to test”. An external transcriber then double-checked the transcriptions and corrected any mistakes (Corden & Sainsbury, 2006). Then, focused coding was conducted as a second cycle coding technique. In this method, major categories are developed from the data (Saldaña, 2015). In a last substage these categories were refined and reorganized according to the literature found related to them. The works by Shettel et al. (1968) and Bitgood (2016) are specially relevant as they were the basis to decide grouping the categories in three major themes: attracting power, holding power and learning power. It’s worth noting that during the categorization a parallel search for relevant literature was conducted, and this literature in turn influenced in a continuous update of the categories. Due to the holistic methodology in this case, a single coding manual is not included in the appendices. Instead, the categories can be found as subsections in Chapter 7 and they are summarized in the following paragraph.

Categories belonging to the theme of attraction power: colour and other stimuli, visibility, transferable attraction, novelty, intergenerational interaction, motivation to read panels, spatial layout of exhibits. Categories belonging to the theme of holding power: enjoyment, comfort, interactivity: property of the exhibit, interactivity: the user as part of the exhibit, social interaction, challenges, testing, diversity of topics and exhibits, linked concepts, lighting, design and maintenance. Categories belonging to the theme of learning power: understanding what science is, inspiration and passion, the science in science centre exhibits, phenomena exposure, immediate apprehend ability, instructive labels, is engagement really a sign of learning?, the telephone game, personal relevance, post-visit engagement.

<u>Inter-coder reliability and the positive decisions ratio</u>

Three researchers coded a set of responses[90] utilizing the coding manuals. Differences in coding were discussed and the manuals were improved iteratively until all researchers agreed (Creswell, 2009; Saldaña, 2015). Since calculating at least one inter-coder reliability index is a best practice (Lombard, Snyder-Duch, & Bracken, 2002), Krippendorff’s Alpha (Krippendorff, 2011) and Average Pairwise of Agreement were calculated with the online tool provided by Freelon (2010). The criterion of good agreement was .7 for alpha (A. F. Hayes & Krippendorff, 2007; Lombard et al., 2002) and 80-90% for pairwise of agreement (Fowler, 2013; M. B. Miles & Huberman, 1994).

[90] Only a few categories were not assessed because the category was not present in the sample (it was added after the rating process). Most of the cases were picked randomly, but a few tricky ones were picked manually to ensure that even the most difficult cases could be sorted.

An issue with Average Agreement is that it inflates the figures (Lombard et al., 2002). But Krippendorff's Alpha does the opposite in some cases. Alpha works fine when codes are present, not when they are not. When most coders do not find a code in the text, but one coder does, alpha plummets. This falsely gives the impression of lack of agreement when, actually, the level of agreement was high (that the code was not present). For this reason, the author created a third index, the Positive Decisions Ratio (PDR):

$$PDR = \frac{TPD}{N \cdot NC}$$

where TPD is the number of times the code was present (positive decisions) according to all coders in total, N is the sample size, and NC is the number of coders.

PDR ranges from 0 to 1 and, to some extent, can be thought of as an effect size of Krippendorff's alpha. For example, when one researcher detected one code once, but the others didn't, alpha was .000. However, what actually happened was that all researchers almost perfectly agreed the category was not present. PDR in that case would be .021, showing the effect size is too small to consider Krippendorff's alpha reliable in that case.Given that PDR is new, it doesn't have small, medium and large values assigned, but the researcher considers .1 as a threshold under which the effect size is negligible.

The minimum value in Average Pairwise Agreement was 88% and 54 of the 64 categories had a Krippendorff's alpha above .7. Nine of the remaining ten had an insignificant PDR (below .1), meaning that their alpha was not reliable. The only category with a low alpha and PDR over 0.1 was one "Other" category, with the reason being that disagreements from other categories accumulated in that category.

Reliabilities were high from the first round of inter-coding[91]. The few disagreements were easily solved by coming back to the definitions and examples of the coding manuals, and by adding extra specifications based on the author's science centre knowledge[92]. It was then decided the author could proceed to code all the data.

---

[91] Do not confuse with the process of creating the Coding Manuals. That required several iterations until they were clear and covered all possible scenarios. This previous work is what lead to high inter-coder reliability from the first inter-coders session.

[92] For example, "Humans are more dangerous to sharks than to people" was thought to be Biology (Bio) by the other researchers, but that was actually mentioned in a planetarium show (Pl).

3.5.3 Descriptive statistics

Descriptive statistics from qualitative data were provided mainly through charts, produced in Excel 365™. Word clouds were produced online with WordItOut™ (Enideo, 2019).

3.6 Summary

This chapter introduced the methods used throughout the research. The chosen approach was mixed methods, wherein quantitative and qualitative methods complement each other to produce a broader picture.

Interviews and focus groups are highly qualitative and were included to complement the main method, surveys. Discovery World and Tūhura visitors were surveyed with four questionnaires on iPads. Each questionnaire had both closed and open questions, producing quantitative and qualitative data simultaneously. Surveys were designed to collect data related to scientific knowledge, science engagement, and self-beliefs in science. Whenever possible, these constructs were assessed before and after visiting the science centre. Some new methods, such as the Visual Discrete Scale, were introduced with a justification of their need and rationale of their construction.

A comprehensive description of qualitative and quantitative analyses was presented, from data pre-processing to inferential testing. A couple of flow diagrams were created by the author to decide whether a set of data can be treated parametrically, and what test is the best for each case.

Chapters 4, 5, and 6 discuss visitor learning from the three main aspects of scientific literacy: scientific knowledge (Chapter 4), self-beliefs in science (Chapter 5) and science engagement (Chapter 6). These three chapters make use of survey data and answer Research Questions 1, 2, and 3. Chapter 7 explores perspectives obtained from staff interviews and visitor focus groups about exhibit characteristics that influence learning, addressing Research Question 4.

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

4.1 Introduction
4.2 Scientific Knowledge
4.3 Non-demographics Learning Factors
4.4 Beyond Light and Electromagnetism
4.5 Conclusions

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 4: SCIENTIFIC KNOWLEDGE: are visitors learning?

4.1 Introduction

One critique of science centres is that visitors simply play rather than learn science. Research questions in this thesis are related to learning at science centres, but learning is a complex concept. In Chapter 1 it was defined as the change in scientific literacy, which, in turn, comprises scientific knowledge, self-beliefs in science, and science engagement.

This chapter examines scientific knowledge, which is the aspect of scientific literacy most typically associated with science learning. The first section discusses the results from assessing learning objectively with a multiple-choice test on science. Visitors to the pre-redevelopment science centre (Discovery World) answered a shorter version of the test. Visitors to the post-redevelopment science centre (Tūhura) answered a longer version of the test. In both cases, the test was answered both before and after their visit. The tests were designed to objectively answer the question of whether science centre visitors take away new scientific knowledge. The effects of age and gender in scientific knowledge and learning are analysed in this section. The following section discusses additional factors that may also influence how much content knowledge visitors learn from their visit.

Sections 2.2 and 2.3 are mainly quantitative and focus on light and electromagnetism, while 2.4 addresses toward the qualitative nature of learning. This section shows the breath and width of knowledge that can be acquired by visiting a science centre.

4.2 Scientific Knowledge

4.2.1 Yes, visitors learn science at a science centre

Scientific knowledge (SK) on light and electromagnetism was tested before and after visiting the science centre with two multiple-choice questions in Discovery World and five in Tūhura. The total score[93] of scientific knowledge (sum of correct answers, not including the control question) before and after visiting Discovery World and Tūhura is shown in Figure 4-1. The increment was significant at both Discovery World scientific

[93] To decide if it would be valid to compare two versus five items, all of them were first tested for significance individually (each item's answers recoded as Correct, Wrong, and IDK groups) using the McNemar-Bowker test dividing. The level of significance in the change from pre to post was similar in all items.

knowledge(N=224, paired t-test t(223)=6.35, $p<.001$, d=0.424, d_{CI}=0.096)[94] and Tūhura (N=456, t(455)=11.9, $p<.001$, d=0.560, d_{CI}=0.068). The change is equivalent to 15% increase in right answers in Discovery World and 13% in Tūhura, but it needs to be acknowledged that it may be partly due to a cueing effect (see section 3.2.2).

The effect size in both science centres falls in what Hattie (2009) catalogues as the 'zone of desired effects learning', i.e., learning that has surpassed what would be expected from formal schooling. Hattie's interpretation of Cohen's d may be of help to realize that informal education may have an similar effect to that of formal education. Although the recurrence in formal education may produce deeper learning than a one off visit to a science centre (see next section), this is more related to the amount of learning time employed in each system, rather than with the efficacy of the system.

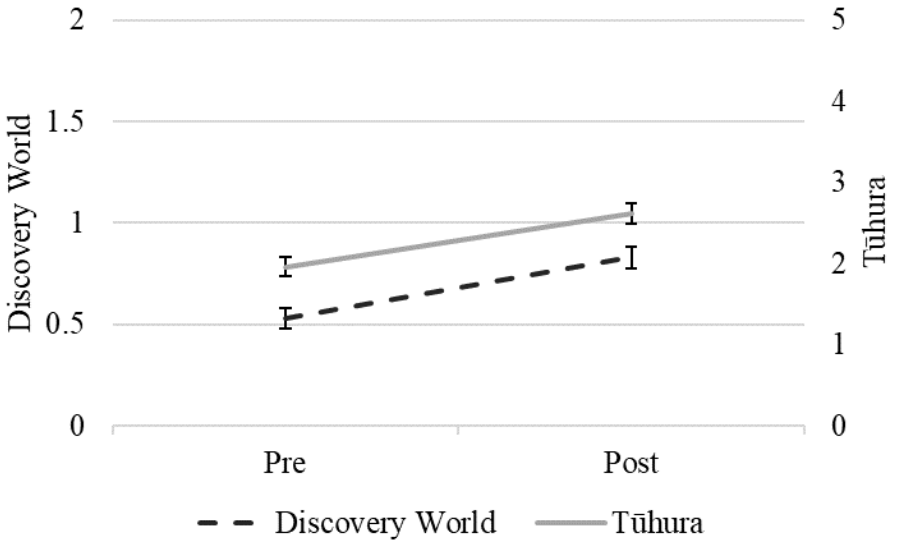


Figure 4-1 Rescaled scientific knowledge in light and electromagnetism before (pre) and after (post) visiting Discovery World (N=224, two items) and Tūhura (N=451, five items). Error bars represent confidence intervals at 95%.

The positive and almost parallel slopes in Figure 4-1 indicate that learning is very similar in size at both science centres. Interestingly, the initial score is considerably higher in Tūhura, which at first glance makes it look like Tūhura visitors are more scientifically literate than Discovery World visitors. To shed some light on the origin of this result, scientific knowledge was controlled for number of visits and questions asked. A partial score was calculated using items SL1, SL2 and SL3, that is, the only items that were asked at both

[94] d_{CI} is the confidence interval of the reported Cohen's d.

science centres before the visit (Table 3-6). Only first-time visitors to either centre[95] were included for this comparison.

Controlling for number of visits and using only questions that were asked before the visit at both science centres, there is no difference in scientific knowledge[96] ($t(238)=0.004$, $p=.997$, $d=0.001$, $d_{CI}=0.094$), as shown in Figure 4-2.

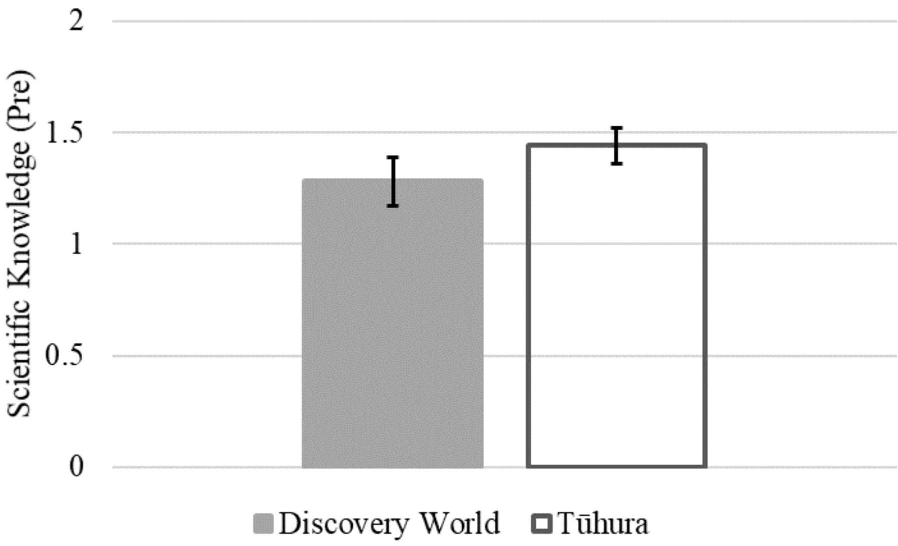


Figure 4-2. Rescaled scientific knowledge in light and electromagnetism before the visit (pre, three items). Only first-time visitors and the three questions in common between Discovery World (n=107) and Tūhura (n=133) are considered. Error bars represent confidence intervals at 95%.

4.2.2 Rome wasn't built in a day

Studying has a large positive effect on long-lasting learning (Karpicke & Roediger, 2008). But over time, conceptual understanding is equally as likely to improve as it is to deteriorate (Falk & Dierking, 2018) and specific knowledge tends to be replaced by the 'big picture' (Falk & Dierking, 2016).

No correlation was found between prior scientific knowledge and the number of visits to Discovery World (Spearman's correlation coefficient $r_s(224)=.012$, $p=.856$) or to Tūhura ($r_s(393)=.035$, $p=.483$). It is likely the frequency with which information is reviewed that influences long-term learning (Yang, Potts, & Shanks, 2018), not the total number of visits. Thus, frequency of visitation may have been a better indicator, as the more frequently a

---

[95] In this case, a first-time Tūhura visitor is defined as not having visited either Tūhura or Discovery World before.

[96] During the analysis above a counter-intuitive collateral result was found: first-time visitors to Discovery World scored higher than Discovery World recurrent visitors. The explanation was simple: there were twice as many children in the group of recurrent visitors (39%) than in first-time visitors (19%).

person visits a science centre, the higher their knowledge and understanding of science (Falk et al., 2016).

If so, the negative result in terms of total number of visits would indicate another thing: informal science learning requires, like anything in life, practice. Ideally, more should be done to invite visitors to come frequently. To encourage this, the Otago Museum has created the Tūhura annual pass. Highly engaged visitors tend to become members (N. Simon, 2010). In the first year of operation, the average visitation for annual passholders was estimated to be 3.8 visits per passholder[97].

Even when frequent visitation is not possible, what sometimes is taken away from a session is dormant knowledge (that another experience will bring to the conscious level) (Krishnamurthi & Rennie, 2012) or has the potential to be learned later (due to the sparked interest) (T. W. Burns et al., 2003; Stocklmayer & Gilbert, 2002).

4.2.3 Visitors also learn science at school

S. Allen (1997) interviewed visitors that interacted with an exhibit 'coloured shadows'[98] to see if they could get more right answers to questions about the nature of those shadows (asked during the interview and later assessed). The success rate in getting the right answers after an intervention was null for visitors under 12 years old, very small for those between 13 to 15 years old, and only considerable for those 16 and above.

The dependence of learning on age is also of interest to this research. A LOESS fit was done on a scientific knowledge scatter plot before and after the visit to Tūhura (Figure 4-3). The independent variable was age. As previously mentioned in section 3.4.2, the LOESS fit does not produce a single regression with correlation coefficients. Instead it allows to see how there is in fact a linear relationship between knowledge and age, only that this relationship is split in two different sections with two different slopes. The domain of one of the relationships is Children and Adolescents, while the domain of the other one is Young Adults and Mature Adults.

[97] Craig Grant, Otago Museum's Director of science engagement, personal communication, 4 October, 2019.
[98] This is the same exhibit as the one on display at Discovery World, and very similar to the one at Tūhura.

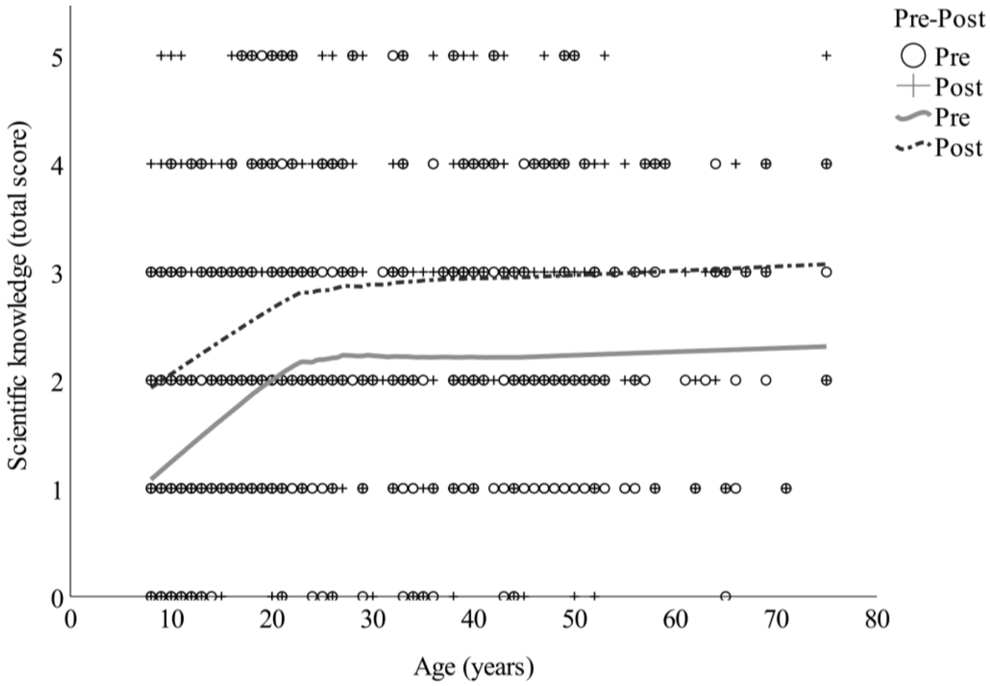


Figure 4-3. Scatter plot with LOESS regressions (α=.70) for scientific knowledge as a function of age, before and after the visit (N=459) at Tūhura.

This result agrees with the conclusion of Ramey-Gassert (1997); both children and adults learn science at science centres. Notice how the increment in scientific knowledge from pre to post does not depend on age. Why S. Allen (1997) got a different result may be due to the nature of the 'coloured shadows' exhibit. How these shadows get their colour is counterintuitive and requires a good deal of abstraction—something that has not yet developed in children under 12 (Piaget, 1968). Also, prior knowledge becomes more important when learning abstract concepts (Krajcik & Sutherland, 2010). This leads to another interesting result from Figure 4-3.

Prior knowledge in Tūhura visitors increases monotonically[99] with age ($r_s(451)=.322$, $p<.001$). Discovery World does not have a LOESS regression because age was not measured as a continuum, but it presented a similar correlation ($r_s(224)=.165$, $p=.013$).

Moreover, notice how there are two well-defined sections. The first one forms a steep slope for school age participants. "From eight to 18 years there is great potential for children and young people to extend their knowledge tremendously… In school, new areas of knowledge continue to be introduced and different skills are required" (Lindon, 1996, p. 50). The influence of school in increasing knowledge about science is not surprising. "Historically, almost all efforts to understand and improve public knowledge of science and technology have begun with the explicit or implicit assumption that formal schooling,

[99] The term refers to a continuous increase, but not necessarily linear. For this reason, Spearman's rho (r_s) is calculated instead of Pearson's r.

particularly elementary and secondary but also post-secondary, provides the vast majority of contribution" (Falk & Needham, 2013, p. 432). Notwithstanding, the parallel upwards shift of curves in Figure 4-3 indicates that the influence of informal learning can be important, even when compared to that of traditional schooling, as also suggested by Falk and Needham (2013).

The subsequent flatter section does not mean adults learn less, but that their priorities tilt their learning to other subjects (Flynn, 2012), not assessed with this instrument (which measured the scholarly subjects of light and electromagnetism). Adults, as opposed to being generalists, tend to become experts in specific domains (Fenichel & Schweingruber, 2010).

What is more important is this: how much people learn[100] from the science centre does not depend on their age (Discovery World: r_s=.034, p=.609, N=224; Tūhura: r=.023, p=.622, N=451). Nobody is too old to stop learning from a visit to a science centre. Indeed, an 83-year-old woman[101] provided the following unsolicited comment after her visit to Discovery World: "Amazing, something I don't see every day, you know, out of the daily cleaning at home and stuff". After she went to the planetarium, she returned to add "I'm getting older, but that still amazes me"[102].

4.2.4 Gender gap in knowledge, not in learning

Females tend to score lower than males in scientific knowledge in some studies (S. Allen, 1997; Kurtz-Costes et al., 2008; Skaalvik & Skaalvik, 2004). In this study, knowledge increased significantly after the visit regardless of gender (Table 4-1). This agrees with (Piraksa, Srisawasdi, & Koul, 2014), who found that gender did not influence scientific reasoning in students in a study in Thailand.

[100] Taken as the subtraction of pre-scientific knowledge from post-scientific knowledge (ΔSK).
[101] She accompanied her husband, with no children.
[102] The researcher wrote down the quotes as soon as possible and to his best knowledge, they are verbatim.

Table 4-1
Changes in scientific knowledge from before to after the visit by gender in Discovery World and Tūhura

Discovery World							
	Males			**Females**			
Pre	M=0.67, CI=0.17	SD=0.79,	Z=3.68, p<.001, dp=30, r=.399, n=85	M=0.44, CI=0.11	SD=0.65,	Z=4.62, p<.001, dp=59, r=.392, n=139	
Post	M=0.94, CI=0.17	SD=0.79,		M=0.76, CI=0.13	SD=0.78,		
Gender difference before the visit: U=5034, p=.035, r=.142							
Tūhura							
	Males			**Females**			
Pre	M=2.23, CI=0.20	SD=1.40,	t(184)=7.13, p<.001, d=0.524, n=185, d_{CI}=0.106	M=1.77, CI=0.14	SD=1.15,	t(266)=9.51, p<.001, d=0.582, n=267, d_{CI}=0.088	
Post	M=2.84, CI=0.20	SD=1.41,		M=2.45, CI=0.16	SD=1.36,		
Gender difference before the visit: t(345)=3.69, p<.001*, n_m=185, n_f=267, d=0.359, d_{CI}=0.096							
Gender difference after the visit: t(450)=3.01, p=.003, n_m=185, n_f=267, d=0.291, N=452, d_{CI}=0.096							

dp stands for the number of discordant pairs.
d_{CI} stands for the confidence interval of Cohen’s d.
*Equal variances not assumed.

Discovery World’s sample size is not large enough to break down gender by age, but Tūhura’s is in most cases. However, age groups’ sample sizes are small for means, so medians are reported instead. The differences in medians from before to after the visit are presented per age group and gender in Figure 4-4. Notice how the gender difference seems to appear until later stages in life, but after visiting the science centre these differences tend to decrease.

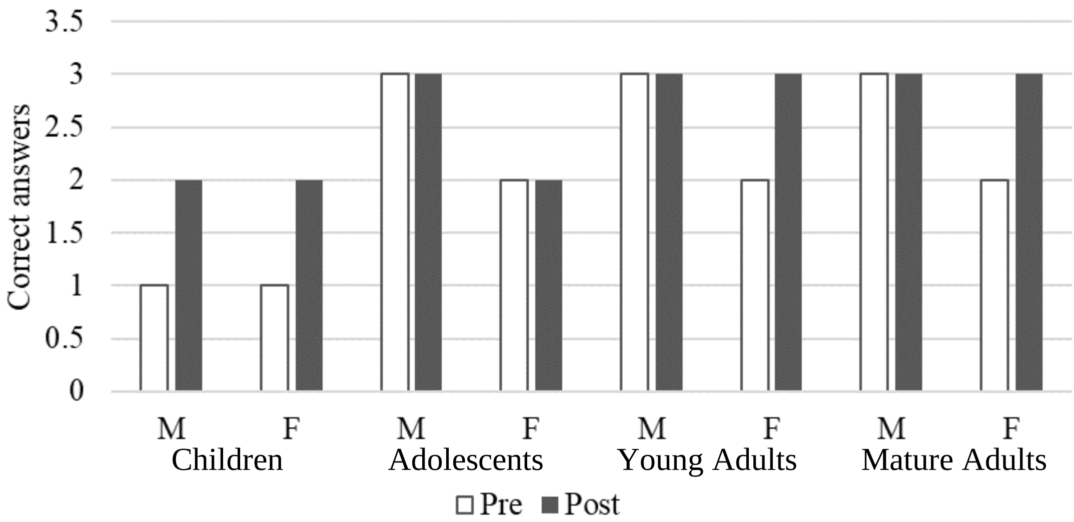


Figure 4-4. Medians of correct answer before (pre) and after (post) visiting Tūhura for male Children (n=56), female Children (n=55), male Adolescents (n=21), female Adolescents (n=55), male Young Adults (n=57), female Young Adults (n=96), male Mature Adults (n=50) and female Mature Adults (n=57). M stands for males and F for females.

To test for significant differences, Table 4-2 shows results from statistical tests comparing the percentage of correct answers (medians) in scientific knowledge before and after the visit per gender and age group in Tūhura. Notice how there is no statistically significant difference between females and males in Children, but there is for Young Adults and Mature Adults. Interestingly, the gap disappears in the latter group after the visit. More research is needed to better understand how learning at science centres indirectly affects the gender gap.

Table 4-2

Statistical significance of differences of correct answers (medians) in scientific knowledge before (B) and after (A) the visit by gender and age group in Tūhura

		Males	Females
Children (8-12)			
Pre-post difference		Z=4.52, p<.001, dp=37, r=.427, n=56	Z=3.52, p<.001, dp=35, r=.336, n=55
Gender difference	B	U=1459, p=.618, r=.052, n=95	
	A	U=1349, p=.252, r=.118, n=95	
Adolescents (13-18)			
Pre-post difference		NA (n=21)	Z=3.73, p<.001, dp=31, r=.356, n=55
Gender difference	B	NA	
	A	NA	
Young Adults (19-40)			
Pre-post difference		Z=3.43, p<.001, dp=35, r=.322, n=57	Z=5.12, p<.001, dp=58, r=.370, n=96
Gender difference	B	U=1887, p=.001, r=.266, n=153	
	A	U=2171, p=.029, r=.176, n=153	
Mature Adults (41+)			
Pre-post difference		Z=2.95, p=.003, dp=30, r=.295, n=50	Z=4.40, p<.001, dp=39, r=.413, n=57
Gender difference	B	U=1013, p=.008, r=.257, n=117	
	A	U=1219, p=.184, r=.129, n=117	

dp stands for the number of discordant pairs.

4.3 Non-demographics Learning Factors

4.3.1 Is engagement a sign of learning?: the role of guidance and self-control

This thesis research follows the constructivist theory of learning by discovery, which considers that learners create and organize knowledge during problem-solving situations (Honomichl & Chen, 2012). This occurs only if there is feedback, worked examples, scaffolding, and elicited explanations; otherwise, unexplained discovery does not necessarily lead to learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Honomichl & Chen, 2012; Mayer, 2004; Tobias & Duffy, 2009).

While children describe science exhibitions as ‘cool’ and ‘fun’, adults appreciate the opportunity to interact with the family and help their children think creatively (Randi Korn & Associates Inc., 2006). When children go in a family group to the science centre, their experience is guided. Parents and other mentoring roles play a critical role in supporting science learning (Fenichel & Schweingruber, 2010). They encourage learning by discovery, utilizing their prior knowledge to help their children direct their learning[103] (Ellenbogen, Luke, & Dierking, 2004; Falk, Dierking, & Foutz, 2007; Falk & Dierking, 2016; Fender & Crowley, 2007; National Research Council, 2009). For example, the author heard a parent explaining to his child how the Earth’s magnetic field was involved in auroras.

In Section 3.3.16 it was mentioned that some data were collected during the science centre sleepovers, and that all attempts to detect learning in young children were fruitless. Only a small group of adolescents showed evidence of learning. During the sleepovers, there were some parents, teachers, and floor staff around, but they didn’t participate in children’s play; they only watched over them. Children interacted with the exhibits under total and unrestricted freedom. It is theorized that the lack of regulation and overstimulation from having the whole place to themselves (only friends their age in a sleepover in a place full of games and a giant slide!) were the factors for why these children didn’t learn as much science as they could have under different conditions.

> When parents take their children to museums, they generally try to make the experience as meaningful as possible for them. They point out key features, read the labels, and engage their child in a conversation about what they are learning. (Fenichel & Schweingruber, 2010, p. 75)

Nevertheless, it is important to acknowledge that a highly social experience which is not focused on science learning is not only valid, but it may bring desirable outcomes in science engagement (which was not measured by the instruments) and social maturation. Hollos and Cowan (1973) and Hollos (1975) studied the behaviour of children in environments with different levels of social and verbal interaction. Conclusions from their studies can be summarized, thusly: “some minimal level of experience in verbal-social interaction appears to be sufficient for the development of logical operations, and a higher threshold is probably required for the development of role-taking skill” (Hollos, 1975, p. 648). In other words, higher social interaction may not be needed to develop scientific logic, but it helps to increase the ability to put oneself in another’s place and consider points of view different from one’s own. This may be invaluable in topics like climate change, where individual actions become global.

[103] Anecdotally, this was seen by the researcher during the surveying time.

The sample size of the adolescents group was too small (N=7) to make any statistically valid inferences. Still, the author would like to propose a couple of reasons that would explain why they learnt more. These explanations are merely speculative and would need to be tested in future research.

First, prior knowledge forms a cornerstone of all subsequent learning, and it is especially important when concepts are abstract (Krajcik & Sutherland, 2010). 'Abstract' is key here. It was already shown in Section 4.2.2 that learning is not related to age, but during the sleepover, children didn't have adults explaining the exhibits to them (e.g., terms like 'electromagnetic wave' require more than a panel to be understood). On the other hand, Year 13 adolescents are just about to enter university, and they already have a basic understanding of pretty much all the concepts needed to get the most out of the exhibits.

Second, emotion regulation has important implications in cognitive development, and it increases with age (McRae et al., 2012; Silvers et al., 2012), until it stabilizes at around 17 years old (Silvers et al., 2012). Year 13 adolescents are no longer chasing their friends down the slide. Their self-control allows them to explore with the calm and critical eye needed to learn.

4.3.2 Is visit time a sign of learning?

As it was shown in Section 3.3.12, time spent at an exhibition can be indication of engagement, but what about learning? The author agrees with Serrell (2010) in that paying attention is a prerequisite for learning; therefore, the amount of time spent in an activity would be a measure of learning, even if the basic idea was already understood. "If the exhibition's message is clear and quickly and easily accessible, high time and high use will not be necessary for understanding the exhibition. But higher time and more thorough use can create more opportunities for visitors to discover more new ideas, to gain a deeper understanding of previously known ideas, or to spend more time contemplating 'old friends'" (Serrell, 1997, p. 121).

Visit time to Discovery World and Tūhura was recorded. Since the aim of this research was to study exhibits, only visitors who interacted with the exhibits were included in the calculation of visit time. Unfortunately, time at the exhibits cannot be isolated from time in the Tropical Forest. Still, there are other factors whose effects can be minimized.

Visitors who attended the Science Show in Discovery World, or the planetarium[104] in Tūhura, were not considered in this calculation.

Under these conditions[105], Discovery World visitors who answered the scientific knowledge test attended for a mean of 56minutes 13seconds (n=124, SD=19m 27s, CI=1m 45s). For Tūhura visitors, the mean was 67m 11s (n=323, SD=23m 19s, CI=1m 18s).

Spearman's correlation[106] between visit time and increase in scientific knowledge (SK after the visit minus SK before the visit) at Discovery World was $r_s(124)=.145$, $p=.108$. At Tūhura, it was $r_s(323)=.105$, $p=.059$. From these results, it cannot be concluded that visit time is a measure of learning. Nonetheless, several factors need to be considered. First, the p-value is not large in either case; actually, it is very close to the significance cut-off of .05 in Tūhura. Second, the visit time is heavily influenced by the Tropical Forest. Third, the test asked questions only about light and electromagnetism, and nothing guarantees visitors that interacted with specifically those exhibits[107].

It can still be concluded that—even though visit time may be a measure of engagement, and engagement may be needed to learn—there is no such thing as a transitive property of quality. Visit time is not necessarily a measure of learning by itself alone. Serrell herself acknowledges the limitations in the notes. "We cannot assume level of interest by time alone, and we cannot assume meaning-making or learning by time alone" (Serrell, 1997, p. 123). Still, considering the small p-values, it is plausible that restricting the data to exclusively consider the time they spent at the Light Zone / Unseen Forces Zone (i.e., the exhibits related to the test) would have resulted in a significant correlation (at least with regular visitors, not sleepover visitors).

4.3.3 Hands-on learning, eyes-on learning

Answers to the scientific knowledge instrument were recoded as Correct, Incorrect and IDK. All of the items (except the control question) were pooled together[108]. Responses

[104] The Planetarium had a different access points within Discovery World, so Discovery World visit times were considered clean from its influence.
[105] Conditions had to be met. If information on any of the variables was missing, the response was not considered in the calculations.
[106] Pearson's r is more 'powerful' than Spearman's rho, as it is parametric, but Spearman was considered more adequate, as it makes more sense that a correlation between visit time and learning is not linear. Pearson's r for Discovery World is $r(124)=.128$ $p=.158$. For Tūhura, $r(323)=.108$, $p=.053$
[107] In Chapter 7 it will be discussed that these exhibits were somewhat hidden.
[108] In this section, n' means the sample size of the available number of responses, not number of respondents (n). For example, n=26 visitors didn't interact with Tūhura exhibits, but since each survey had 5 items, there were 130 possible responses. n'=127 because three respondents skipped one item each.

from visitors who interacted with the exhibits were kept in one group and responses from those who did not were kept in a different group.

Tūhura visitors who interacted with exhibits changed their answers significantly (McNemar-Bowker test $\chi^2(3,n'=1973)=166$, $p_{asym}<.001$, DPRS=14.0). The non-interacting group didn't ($\chi^2(3,n'=127)=3.628$, $p_{asym}=.305$, DRPS=0.007). Figure 4-5 shows this graphically. Notice how the amount of changed answers was the same in both groups (33%). However, those who interacted with the exhibits have a large net flow towards the right answer, while the distribution of those who did not interact is more random.

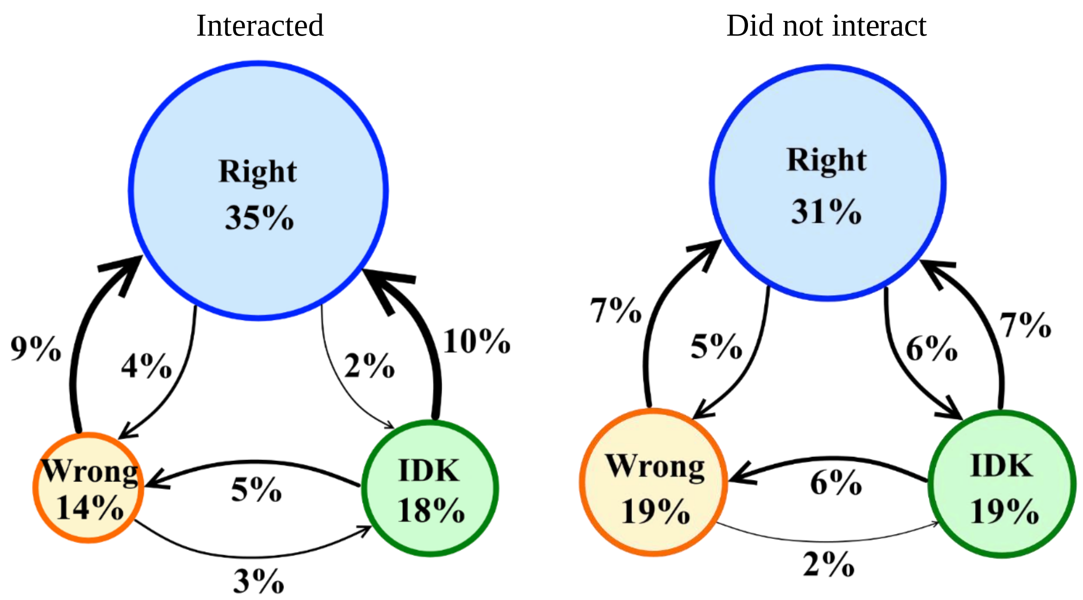


Figure 4-5 Learning flow diagrams for Tūhura visitors who interacted with the exhibits (left, n=409, n'=1913) and those who did not (right, n=26, n'=127). n stands for the number of respondents, n' for total number of responses.

Discovery World visitors who interacted with exhibits also had significantly different responses to knowledge questions after their visit ($\chi^2(3,n'=358)=36.604$, $p_{asym}<0.001$, DPRS=0.102). The non-interacting group, as before, didn't change significantly ($\chi^2(3,n'=90)=6.455$, $p_{asym}=0.091$, DPRS=0.072). However, it is worth noting that the p-value of the latter is close to the significance cut-off.

To the researcher's best knowledge, there is no post-hoc test appropriate for the McNemar-Bowker test (except doing manual McNemar 2x2 tests). Instead, the learning flow diagram (Figure 4-6) shows how visitors' answers moved from before to after the visit. Notice how both diagrams are very similar. This points towards an interesting conclusion: non-interacting Discovery World visitors were also learning. A larger sample might have been able to prove it statistically.

If so, the question changes from ‘if’ to ‘why’ they learnt. If we consider the possibility of visit time being an indirect measure of learning (Borun, 1998; Serrell, 1997, 2010), then that could shed some light. Interacting visitors at Tūhura stayed significantly longer than non-interacting visitors ($t(742)=3.54$, $p<.001$, $n_{NI}=52$, $n_I=692$, $d=0.516$, $d_{CI}=0.144$). This was not so at Discovery World, where the difference was not significant ($t(51.9)=1.21$, $p=.232$, $n_{NI}=45$, $n_I=179$, $d=0.239$, $d_{CI}=0.167$). See Section 6.2.1 for more information about visit time.

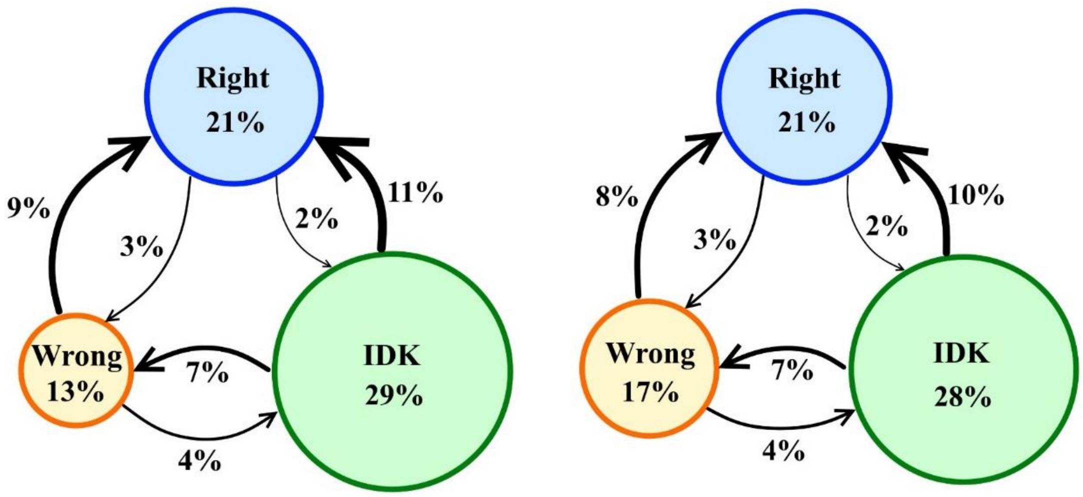


Figure 4-6 Learning flow diagrams for Discovery World visitors who interacted with the exhibits (left, n=179, n’=358) and those who did not (right, n=45, n’=90). n stands for the number of respondents, n’ for total number of responses.

The interpretation is that even when these visitors were not interacting physically with the exhibits, they were still interacting visually. It is possible to learn by observation (Falk & Dierking, 2016), as watching others interact with the exhibits is also an active behaviour (Randi Korn & Associates Inc., 2006; Barriault & Pearson, 2010). “As visitors watch others, they may learn how to use the exhibit or simple enjoy mentally engaging with the activity before physically doing so” (Randi Korn & Associates Inc., 2006, p. 11). This effect may have been bigger in Discover World because the *Plasma Room* (part of the Light Zone) was especially engaging and completely visual. Given that not many exhibits were as engaging as the *Plasma Room*, it didn’t have much competition. On the other hand, Tūhura is much more varied. It is plausible that a test based on non-visual topics would have led to no learning in non-interacting visitors at both science centres.

4.3.4 The role of panels

Visitors tend to fit into two groups: those who read most panels and those who read few or none at all (Falk & Dierking, 2016). It is a big challenge for science centres to design clear and enticing panels that enhance the experience but also, if not read, don't preclude a valuable visitor experience. Not even diligent visitors read everything (Serrell, 2015).

Change in scientific knowledge was examined for Tūhura visitors who interacted with the exhibits, distinguishing whether they read some panels or not[109]. The difference in learning (ΔSK) was not statistically significant (Table 4-3), meaning that those who did not read the panels learned just as much as those who did. How closely related were the knowledge questions with the panels content varies. For example, 'SL2. Can atoms give off light?' (Table 3-6) is not mentioned explicitly in any panel, while the answer to 'SL4. What colours can you combine to form white?' appears directly stated on *Coloured Shadows* panel.

Table 4-3
Scientific knowledge before (pre) and after (post) a visit to Tūhura depending on whether visitors read some panels or not

		M	SD	CI	ΔSK	Test on ΔSK
Non-readers (n=115)	Pre	1.71	1.21	0.22	0.60	$t(425)=0.544$, $p=.587$, $n_{NR}=115$, $n_{PR}=312$, $d=0.061$, $d_{CI}=0.109$
	Post	2.31	1.39	0.25		
Panel readers (n=312)	Pre	2.08	1.31	0.15	0.67	
	Post	2.75	1.39	0.15		

ΔSK stands for the change in scientific knowledge, defined as post-knowledge minus pre-knowledge.

Further research is needed to explore possible reasons for this result. Perhaps panels enhance the experience in a way the scientific knowledge instrument cannot detect; panels might simply not have any influence on visitors, or the self-explanatory nature of the exhibits made the panels redundant. A discussion on desirable panel characteristics is presented in Chapter 7.

4.4 Beyond Light and Electromagnetism

The previous sections dealt with quantitative results from visit time and the scientific knowledge instrument. But that instrument was designed specifically for light and electromagnetism, and a science centre is a lot more. While it is not possible to include all possible topics in a test, qualitative results can show how extensive learning actually is.

[109] Those that didn't interact were only 26, from which 21 didn't read panels and 5 did. Too small subgroups to do stats with them.

4.4.1 Self-reported learning

After a visit to Discovery World, a young respondent, around 11 years old, mentioned "I learnt a lot in there" when she eagerly came back to fill out the second part of the survey. She carefully read each question, taking her time. After her father repeated several times, "If you don't know, click I don't know", she replied, "It's not that I don't know, I just need to be sure". Then she explained that she wanted to recall it right, because there were many things she learnt. After the explanation, her parent seemed surprised and stepped back in a clear move to give her space while encouraging her to "think well".

The anecdote above paints a scenario where visitors were aware of how far their knowledge can reach. Only 36% (356) of Tūhura visitors specifically said that they came to the science centre to learn some science in the pre-visit survey. However, when they were directly asked, after their visit, whether they had learnt something they didn't know before, 78% (N=354) of Tūhura visitors agreed.

In agreement with previous results, the capacity to learn doesn't depend on age, as there was no age-wise pattern (Children, 77%, n=91; Adolescents, 68%, n=74; Young Adults, 81%, n=113; Adults, 85%, n=71). Interestingly, the percentage of females reporting new learning (82%, n=213) was significantly higher ($\chi^2(1)=6.37$, p=.012) than that of males (72%, n=138). How this fits with the gender gap mentioned above is not yet clear. It may be related to the Tropical Forest, as it is well-known that females are more attracted to biological science, while males to physical sciences (Akarsu & Kariper, 2013).

In the selected set of Modes of Learning Inventory (MOLI) items used in this study's surveys, 86% (N=198) of visitors reported high or very high learning[110]. Once again, there is no clear age pattern (Children, n=59, Mdn=3.83, IQR=0.83, CI=0.33; Adolescents, n=37, Mdn=4.17, IQR=0.75, CI=0.17; Young Adults, n=42, Mdn=4.00, IQR=0.71, CI=0.17; Adults, n=58, Mdn=4.17, IQR=0.67, CI=0.17). Using these items, answers from males (n=78, Mdn=4.08, IQR=0.83, CI=0.08) and females (n=117, Mdn=4.00, IQR=0.83, CI=0.17) were not statistically different (Mann-Whitney U=4342, p=.564, r=.041).A difference between the MOLI and the direct question is that MOLI does not restrict the response to "new" learning. That may explain why MOLI reports higher learning; remembering something we have forgotten can also be considered learning (Falk & Dierking, 2016). In that sense, these two methods complement each other.

[110] Results were recoded as Very Low (6 to 10 points), Low (11 to 15), Medium (16 to 20), High (21 to 25) and Very High (26 to 30). Descriptives were rescaled to values from 1 to 5.

Learning is an individual process, exemplified by the following two comments in response to the question ‘Can you give an example of something that you learnt?’: “That you can balance an object on the to[rqu]e board if you get the object to have a matched to[rqu]e” (T, F, 19). “That if you spin the ball in the opposite direction that the disc is spinning, it stays on there longer” (T, F, 52). These two visitors caught what the *Torque Table* exhibit was trying to convey. The response of the former appears more conceptual, and she is using the terminology learnt at the panel. The second visitor’s explanation is practical and direct, and her learning may have occurred primarily by hands-on experimentation rather than reading the panel.

4.4.2 What was cool to learn about?

Self-reports of learning may raise questions about accuracy, but they also allow for a broader and deeper picture (Falk & Dierking, 2016). Visitor comments from surveys made it clear that the science centre offered a lot more than light and electromagnetism. “[I] think you cou[l]d come lots and learn something new every time” (F, 50, Tūhura). To look at how broad this learning is, open responses to complete the statement, “It was cool to learn about…”[111], were categorised by topic in both science centres (Figure 4-7).

[111] “Have your say! Let us know what you liked or didn’t like of Tūhura’s exhibits or how they can be improved” and “Can you give an example of something you learnt?” were included in Tūhura’s code by topic (see Chapter 3).

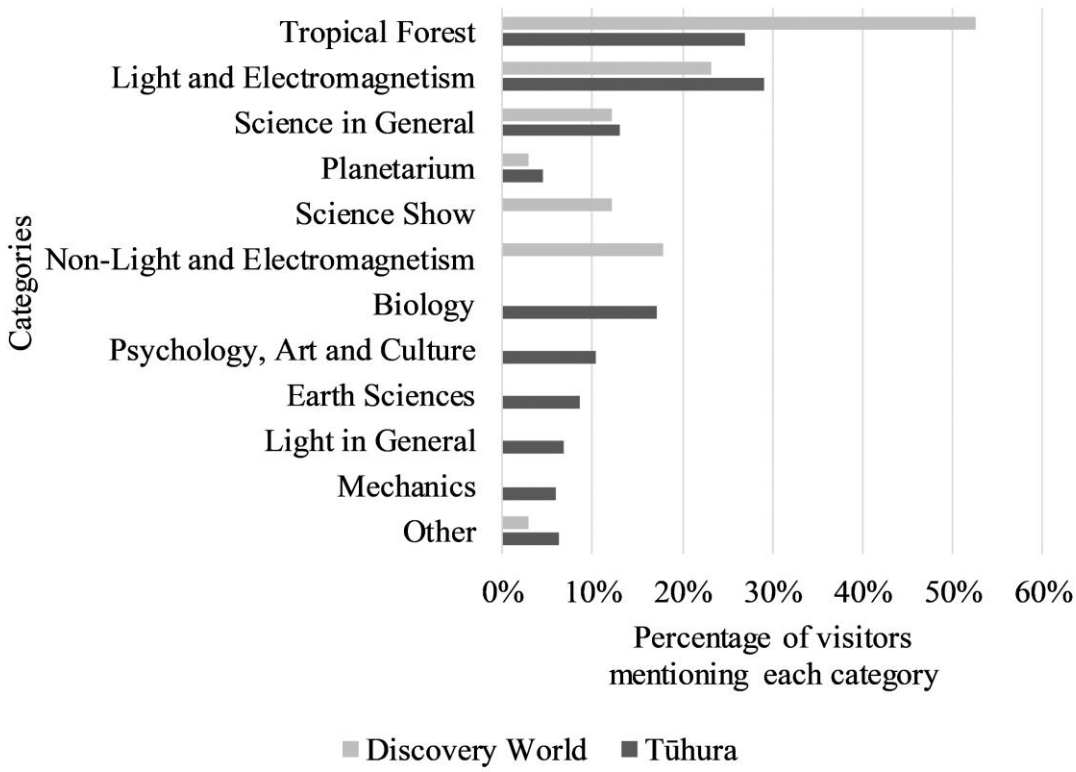


Figure 4-7. Topic categories of visitor responses in completion of the survey statement: "It was cool learning about..." Percentages were calculated by category mentions (CM) with respect to the number of respondents (NR)[112] (Discovery World: N(CM)=214, N(NR)=173; Tūhura: N(CM)=784 , N(NR)=610).

The Tropical Forest clearly dominated visitor attention at Discovery World, as can be seen in this quote: "The butterflies and their habitat that is mostly what we came for on our short visit" (DW, F, 19-40). Light and Electromagnetism was the other popular topic in Discovery World. "Lights, and the way you can use for example electricity and I also thought it was cool seeing all the different things you can do" (DW, F, 8-12).

Tūhura presented broader and more evenly-distributed responses. Light and electromagnetism surpassed the Tropical Forest as the most cited topic. "I learnt [that] my sister has rea[l]ly small hands and that my nose is cold [I] could see this thr[o]ugh the inf[r]ared camera and [I] though t[h]at it was real[l]y cool" (T, F, 9). "How plasma is a state and when you touch it electrons go through your finger creating a small electrical charge [I] think" (T, F, 16). "Monochromatic light is a trip mannnnnnn!!" (T, M, 23). Tropical Forest followed very close. "That the butterflies come from the [P]hilippines. That they survive between a week and a month" (T, F, 44). "That there is only one breed of butterfly that is allowed to breed in the exhibit. It is the owl butterfly" (T, F, 11).

[112] Notice that this implies that the sum of all can surpass 100%.

Other areas with strong presence were the following:

- Biology
    - "My mind was BLOWN by the living chicken foetus [*Chicken Embryo*[113]]" (T, F, 49).
    - "how your leg bones are the hard working ones when you are riding a bike" (T, M, 10).
    - "They staple people's ribs back together after heart surgery" (T, F, 17).
- Earth science
    - "Watching the live seismometer was amazing. Fascinating to see how the earth is active" (T, M, 33).
    - "Water is very strong, it requires a lot of knowledge and understanding how to build a dam" (T, F, 65).
- Psychology
    - "When touching the *Chicken Wire* your hands feel velvety. Because your brain is tricking you" (T, F, 11).
- Mechanics
    - "How the different shapes created different flights over the air vent" (T, F, 21).

Most people tend to attribute learning to the place where they first acquired the information, but learning is also about strengthening existing knowledge (Falk & Dierking, 2016). At least some visitors in this study recognise that. "Plasma the fourth form of matter w[a]s something I knew but [I] almost forgot previously" (T, F, 33). "Recalling torque and inertia was [a] lea[rn]ing event - need to go back to my physics texts of 40 years ago!" (T, M, 58). The museum can take advantage of what a visitor already 'sort of knew'. These quotes are evidence that formal and informal education can work together to help people learn and consolidate their learning.

But the museum goes beyond refreshing memories, as illustrated by this touching quote. "Any exhibits involving light interest me as [I] work with people with dementia and providing these people with different light experiences can potentially 'unlock' parts of their brains" (T, F, 55).

A redevelopment goal was to make the science centre a place where younger children not only have fun, but learn science. Any doubt of whether children learn at Tūhura should

[113] The name of this exhibit is technically Growth and Development, but Chicken Embryo is so popular among visitors and staff that the latter is the name it will be referred to throughout the thesis.

be settled by the following response from a young visitor: "1. I have learned how to make still objects move at the animation station 2. Through an experiment I have learned how humans conduct electricity 3. I learned that white has many different colours" (T, F, 9).

The first step that leads to significant learning is exploring, playing around; the second one is questioning (S. Allen & Gutwill, 2009). The effect of the science centre does not stop with learning factual science—adults and children alike develop the sense of inquiry associated to science and scientists. "How you could create white by using the colours "Red+Blue+Green=[W]hite"[.] I wonder if [I] could make white using paints?" (T, F, 11)[114].

> How optics work in the human eye and how they are used in photography. What the science is behind everyday objects that we take for granted. For example is it a similar process working to make this touch screen respond as the electromagnetic examples in the *Plasma Tubes*? (DW, F, 41+)

4.4.3 What would be cool to learn about?

Learning flows easier if there is some prior knowledge and the topic resonates with the visitor (Falk & Dierking, 2016; Gee, 2012; Krajcik & Sutherland, 2010; Lave, 1988; Lave & Wenger, 1991; Mattar, 2018). It is impossible to cover all the areas of science a visitor might have knowledge of or interest in; but, to find out if there were any big topics missing in the science centre, visitors survey responses to "It would be cool if Discovery World / Tūhura had an exhibit about..." were categorised by topic (Figure 4-8).

[114] Scientific inquiry is a desired outcome, but it can lead to misinterpretations if not correctly guided. For example, in this case it would be enough this girl tries to make white paint by adding red, green and blue paints. Since mixing lights is an additive phenomenon, but mixing paints is subtractive, she would obtain a dark mix, not white, which is a completely counterintuitive and even confusing result. See sections 7.4.7 and 7.5.8 to find more on the importance of testing variations and undesired collateral misunderstandings.

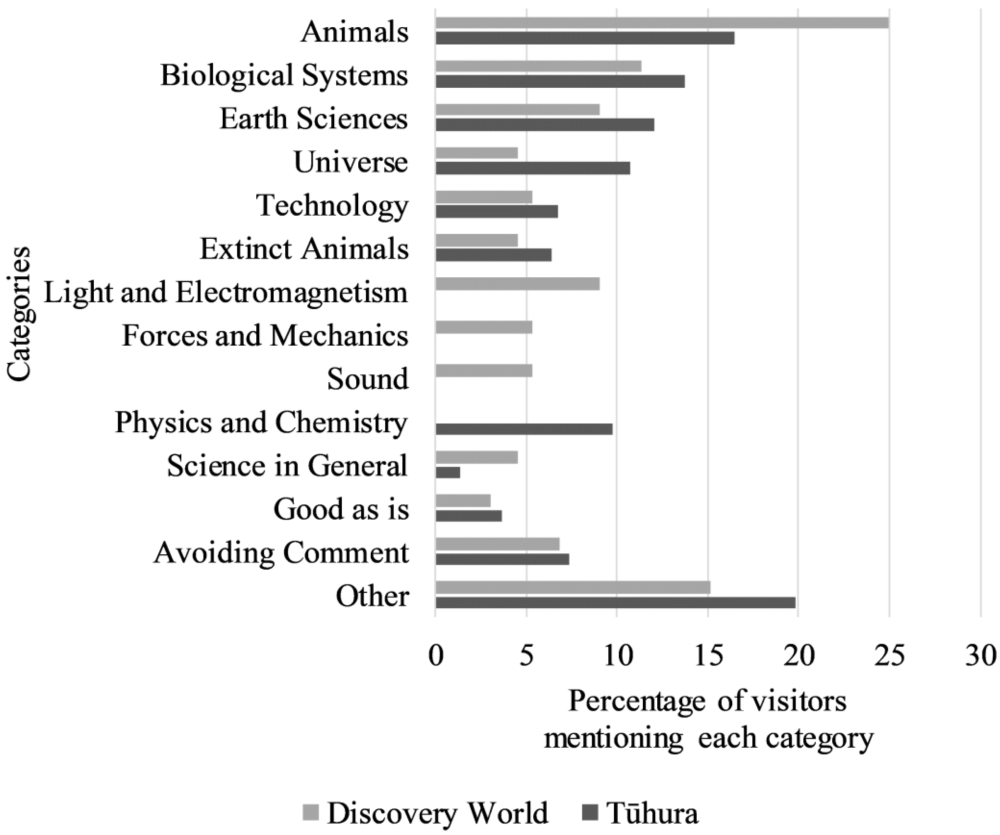


Figure 4-8. Topic categories of visitor responses in completion of the survey statement: "It would be cool if [Discovery World / Tūhura] had an exhibit about...". Percentages were calculated by category mentions (CM) with respect to the number of respondents (NR) (Discovery World: N(CM)=144, N(NR)=132; Tūhura: N(CM)=322, N(NR)=298).

Several things stand out in Figure 4-8. First, Animals were heavily requested (even more, including Extinct Animals) at Discovery World. This topic was popular with children ("[Dinosaurs may be] or birds" (DW, M, 8-12)) and adults alike ("Perhaps something involving more animal science like the *Dung Beetle* exhibit" (DW, F, 19-40)). This is not surprising; more than 700 million people visit zoos and aquariums every year (World Association of Zoos and Aquariums, 2018), positioning them as key forum for science engagement and learning (Wagoner & Jensen, 2010; Waller, Peirce, Mitchell, & Micheletta, 2012).

The decrease in desire to see more exhibits related to Animals in Tūhura agrees with the fact that Tūhura has more exhibits related to them. Still, there is room for improvement. "A little more detail on taxonomy would be cool. We had a go at sorting the creatures but got a bit stuck on trilobites and the one that starts with 'm'. We didn't know what defined the creatures from those categories" (F, 32).

The categories of Light and Electromagnetism, Forces and Mechanics, and Sound decreased so much that they were not considered an independent category in Tūhura, anymore. Like with Animals, this indicates a better coverage by Tūhura of those topics.

Biological Systems and Earth Sciences were two categories that Discovery World adult visitors were particularly keen to see more of. “Memory genetics animal adaptations” (DW, F, 41+). “Soil development and erosion [a]ffecting world agriculture and food supply” (DW, F, 19-40). They were still asked for in Tūhura. “Human-bacterial interaction” (T, M, 26). “Seasons and mahi[ng]a kai”[115] (T, F, 65).

Within the category of Technology, transport was an especially popular topic. “Something about planes or streamlining cars / vehicles” (T, M, 23).

There was a substantial increase in requests for more exhibits related to the Universe. A potential reason for why mentions doubled is that, unlike in Discovery World, the planetarium now shares an entrance with Tūhura. Despite having different tickets, it is possible that more planetarium visitors were also going to Tūhura and want to see more of the same topic in the science centre.

The appearance of Chemistry as a topic that Tūhura visitors asked for may be partially due to the disappearance of the Science Show. Among its topics, it used to cover chemistry. “Learn about making dry ice” (DW, F, 8-12). Science communicator-visitor interaction is also mentioned as something they miss. “Doing exper[i]ments live with the public” (T, M, 9).

An overarching theme behind several requests in Tūhura was making information regionally meaningful. “Australian animals” (T, F, 13). “Dunedin landscape for example volcanic history. Dunedin weather” (T, F, 51).

4.4.4 Proto-scientists

A total of 356 slips from Tūhura’s Questions & Answers board were read from July to October 2018. They were divided into three categories[116]: Tūhura and its exhibits (134), Tropical Forest (65), and Other (157). These visitor questions and comments demonstrate that Tūhura fosters deep inquiry: “how does the human brain think?”, “what would it be like

[115] In Māori culture, mahinga kai is about harvesting food and appreciating the value of food and other natural resources.

[116] Science communicators were in charge of saving discarded posts, but due to high turnover that was not clear to all, and most of the Q&A slips were not saved. Others saved only the comments, not the questions. Some slips were collected by the author, by taking pictures on random days. Therefore, the quotes should be taken as examples of what some visitors ask/comment, not as a statistically representative sample.

if the sun was gone?", "how do astronauts stay the right temp[e]rature in space?", "what kind of petrol do rockets need?".

Their comments and questions also complement section 4.4.3 by showing what topics would be worth addressing in the future. "Why do people believe that the earth is flat?". "exhibit idea what a shadow looks like in a solar eclipse".Other enquiries were directly related to exhibits currently on display. These exhibits could be used to address them. "how do hearts grow inside the yo[l]k of the egg?". "What is a Photon?". "What is the temperature of the core of earth?". "I loved the s[c]ience center. It was amasing! how did the skeleton move at the same time as [u]s? my favorite bit was the bumpy colourful slide" (Age 7). "What came before the big bang?". "Why did the green light bounce of[f] the torna[d]o but you could still see it on the torna[d]o".

4.4.5 QuEST for Science

Tūhura doesn't have a physical space dedicated to special long shows, as Discovery World had for the Science Show. Nonetheless, science communicators occasionally provide short demonstrations using the exhibits. There is also now a free science show, called QuEST for Science, that runs in various places throughout the museum, outside of Tūhura. Anecdotally, it was noticed that Tūhura visitors, in general, didn't leave for the QuEST for Science show and come back—probably because making visitors go outside for something is an unconscious invitation to finish the visit (Falk & Dierking, 2016).

Since the show is not part of the science centre anymore, it was not included in this study. Nonetheless, it was deemed important to mention anecdotal observations by the author about its effect on the visitors. In a show, the science communicator asked the public if they knew something that was really, really black, the blackest thing of all (she was going to talk about Vantablack, a coating made of carbon nanotubes that does not emit light, only absorb). Most answers were as expected, 'black', 'the sky at night'. But a boy around 6 or 7 years old boy replied: 'When there is nothing'. That is the best definition of 'black' (absence of light) the author has ever heard.QuEST for Science may not be part of Tūhura, but providing a free show that anyone can attend is an excellent opportunity to engage visitors with science outside of the science centre.

4.5 Conclusions

This chapter focused on the fundamental question of whether a visit to a science centre results in science learning. To investigate it, a scientific knowledge set of multiple-choice questions on light and electromagnetism was asked to visitors of Discovery World and Tūhura, before and after their visit. Results indicate that visitors indeed learnt science. However, it's important to acknowledge that cueing (having the same set of questions before and after with the same group of respondents) could have influenced the results.

A common view is that learning cannot be measured with 'traditional' tests in informal settings, because of a possible alienation and the breath of informal learning (A. J. Friedman, 2008; National Research Council, 2009). Both concerns are perfectly valid, and this research acknowledges them.

Learning is a complex matter, and its breath was acknowledged by studying it from a mixed methods approach. For example, a quantitative test showed a similarly significant increment of scientific knowledge in both science centres. But it was the qualitative nature of the open questions what revealed that the learning spectrum in Tūhura is broader than it was in Discovery World. Open questions not only showed the variety of topics visitors said they learnt within, but also produced evidence of visitors being able to take what they experienced at the science centre and extrapolate it to personally-relevant fields.

Care was taken to decrease any possible risk of alienation. The methodology around the survey was designed to give visitors a non-threatening environment. As an example, surveys were kept as short as possible. The longest median for any pre or post survey was 4:04 minutes (see Appendix D). Visitors didn't report any discomfort, nor it was detected by the author.

Self-reporting has been widely used as a measure of learning, *assuming* that it is a reliable reflection of actual learning. In regard to RQ1, this research finally demonstrates that, yes—provided all needs and cares are met, a multiple-choice test can reliably measure scientific knowledge learning at science centres.

The moderating variables, age and gender, turned out to be important factors in prior knowledge, although in different ways. scientific knowledge grows quickly during childhood and adolescence, but it reaches a plateau in adulthood. On the other hand, gender doesn't play a role in young children, but adult females in this study showed significantly lower scientific knowledge than males.

How scientific knowledge varies with age seems to depend directly on schooling, showing the importance of formal education in people's knowledge. Knowledge not only stays after people move on from school, but keeps growing afterwards. The gender gap will be further discussed in Chapters 5 and 6. Meanwhile, an important result deserves attention, as it partially answers RQ3: even though prior knowledge depends on age and gender, changes in scientific knowledge are not influenced by these variables. Any visitor—male or female, young or old—can take away a significant amount of scientific knowledge from visiting a science centre.

Aside from evidence of science learning, two other important questions were studied. It is long been assumed that engagement and visit time can be used as signs of learning. What was found here does not sustain both claims as separate factors, but there is evidence indicating that a combination of both may indeed produce indirect evidence of learning. Prior knowledge enables substantive learning (Feher, 1990), but so does engagement (Grinell, 1988), life experience (Bulunuz & Jarrett, 2010), and personal relevance (Rosenthal, 2018). These other factors are related to self-beliefs in science and science engagement. They will be discussed in Chapters 5 and 6, respectively.

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

5.1 Introduction
5.2 Self-efficacy in Science: I can do it
5.3 Scientific fluency: I know that
5.4 Self-concept in Science: I'm good at this
5.5 Why the Likert Scale Didn't Detect a Change
5.6 Gender Differences in Self-beliefs in Science
5.7 Towards the Validations of Self-reporting to Assess Knowledge

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 5: SELF-BELIEFS IN SCIENCE: to know that you know

5.1 Introduction

The previous chapter was dedicated to one aspect of scientific literacy: scientific knowledge. Change in knowledge was assessed objectively with a pre-post visit multiple-choice test. Results clearly showed that it was possible to reliably assess scientific knowledge with this instrument. Moreover, it demonstrated that visitors learn science by visiting a science centre, and that this learning is independent of age or gender.

This chapter explores a different aspect of scientific literacy: self-beliefs in science. Analysing the opinion of one's self regarding science is a complex matter that needs to be approached using several constructs. The first one is self-efficacy in science, a self-belief related to the confidence in one's own scientific capabilities in specific areas.

The second construct, scientific fluency, is a new one in the field of science communication. Its corresponding instrument was developed with the aim of assessing self-beliefs of one's scientific knowledge. If the self-perception is accurate and the instrument to measure scientific fluency is reliable, then it should mirror scientific knowledge to some extent.

The third self-belief construct is self-concept in science. It is similar to self-efficacy in science, but instead of confidence in specific capabilities, it is related to the general perception individuals have about themselves with respect to science. Self-concept was also used to evaluate the performance of the Visual Discrete Scale, a newly developed alternative to Likert-type scales. The new instrument was designed under principles expected to alleviate some of the deficiencies in the Likert format. Results from assessing self-concept in science with both a Likert-type scale and a Visual Discrete Scale are discussed and compared. At the end of the chapter, the relation between self-beliefs in science with scientific knowledge is discussed.

5.2 Self-efficacy in Science: I can do it

Self-efficacy in science is the perceived capacity for doing specific science-related tasks in given situations (Table 1-3). It measures an individual's confidence in performing science-related tasks (Bandura, 1986; Bong & Skaalvik, 2003; OECD, 2009, 2018). Self-efficacy is heavily influenced by personal experiences (Jansen et al., 2015) and strongly linked to previous achievement (Bandura, 1986; Bandura & Locke, 2003; Diseth et al., 2014; Honicke & Broadbent, 2016; Komarraju & Nadler, 2013; Valentine et al., 2004; Zajacova et al., 2005; Zeldin et al., 2008).

The 5-item scale used to measure self-efficacy in science[117] in Tūhura (Table 3-4) showed a significant increase ($t(226)=17.5$, $p<.001$, $d=1.16$, $d_{CI}=0.101$) from before the visit ($M=2.57$, $SD=1.13$, $CI=0.15$) to after the visit ($M=3.56$, $SD=1.13$, $CI=0.15$). Visitors were asked to rate their confidence about light and electromagnetism topics on a scale from 1 to 5, with 1 being 'Not at all confident' and 5 'Very confident'.

Scores were rescaled[118] and the 21 possible scores were condensed in four groups[119] in Figure 5-1. It can be seen that pre-visit responses accumulate on the low confidence left-hand side of the scale and post-visit responses on the high confidence right-hand side. Only 24% of respondents were moderately or very confident before the visit, but this increased to 53% afterwards.

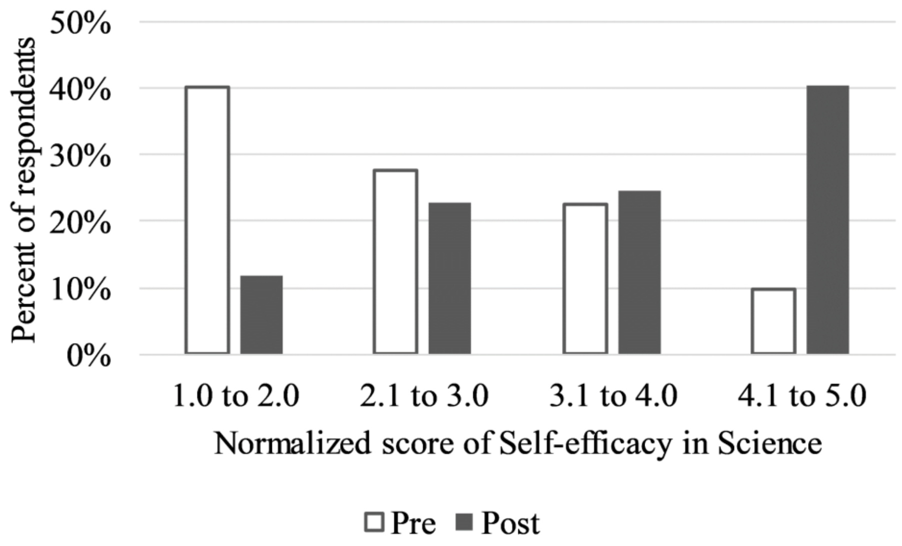


Figure 5-1. Self-efficacy in science before (pre) and after (post) a visit. Score has been rescaled from 1 to 5 and grouped in four brackets for ease of interpretation.

This behaviour seems to be due to the nature of self-efficacy. Since self-efficacy is related to capacity, visitors tended to dichotomize their answers in *cannot* (lower end of the

[117] See Chapter 2

[118] The total score's range is [5, 25]. The rescaled score is obtained by dividing the total score by 5.

[119] Recoding results as Not at all confident (5 to 8 points), Little confident (9 to 12), Somewhat confident (13 to 17), Moderately confident (18 to 21) and Very confident (22 to 25). Descriptives were rescaled to values from 1 to 5.

scale) and *can* (upper end), producing a big jump in scores when changing from *cannot* to *can*. This behaviour might have been exacerbated by the topics under assessment (light and electromagnetism), as terms like 'infrared light' may be foreign to what is heard in daily life, resulting in answers on the *cannot* side before the visit.

It is difficult to assess the level of agreement of this result with previous literature because, as seen in Chapter 1, the few studies in self-efficacy before and after visiting a science centre have their understanding of self-efficacy not well-defined (e.g. Martin et al., 2016) or do not report the items used (e.g. Sasson, 2014). However, the result agrees with that of Bong and Skaalvik (2003) in that self-efficacy is malleable and, therefore, it was expected that visiting a science centre would increase visitor confidence in performing specific tasks related to science.

Self-efficacy in science is not related to age (N=223, r_s=.114, p=.086), as assessed before the visit. However, the p-value is close to α and after the visit age becomes a significant factor (N=223, r_s=.222, p=.001). Therefore, age is a variable to consider in self-efficacy, but is not as important as gender, as will be seen in Section 6.4.

5.3 Scientific Fluency: I know that

5.3.1 Scientific fluency and scientific familiarity

Fluency is the ease with which an individual can extract information from the cognitive system (Kellman et al., 2010; Shimamura & Palmer, 2012; J. K. Smith, 2014; L. F. Smith & Smith, 2006; Tsai & Thomas, 2011). Scientific fluency is the perceived science-related knowledge that facilitates comprehending natural world phenomena (Table 1-3). It is an adaptation of aesthetic fluency (J. K. Smith, 2014; L. F. Smith & Smith, 2006) to measure perceived science-related knowledge. Scientific Familiarity is used here to be the perceived knowledge about a specific science term or scientist (Table 1-3). The instrument for scientific fluency consists of items that measure familiarity with five scientific concepts and five scientists (see Section 1.3.11). Prior familiarity with each of the concepts and scientists was ordered and split by cohort (Figure 5-2). Children have the lowest familiarity with all items, which hints that scientific fluency might indeed be measuring something similar to what was measured with the scientific knowledge instrument.

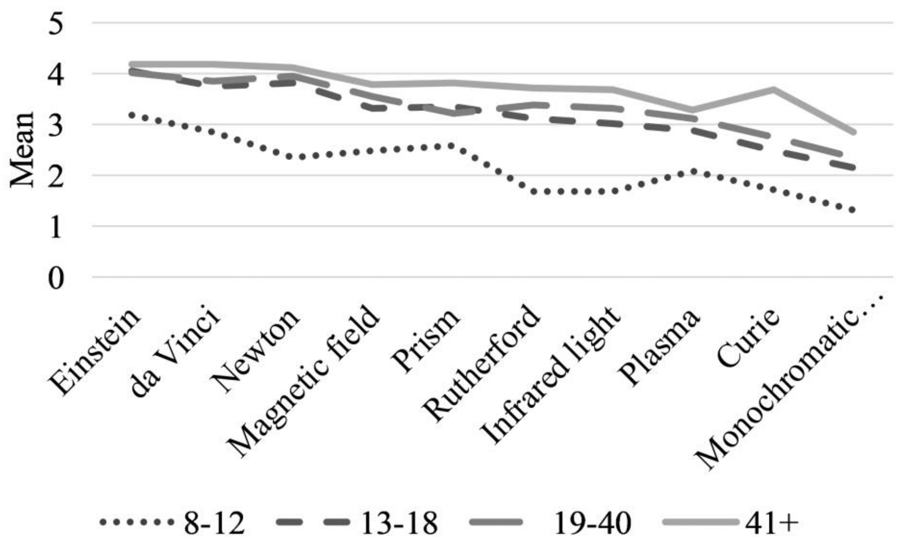


Figure 5-2. Relation between age and prior familiarity. Scientists and concepts are ordered from the highest overall mean to the lowest.

The peaks in Prism and Plasma in Children can be explained by the full primary school curriculum (years 1-8, ages 5-13), which includes light and electromagnetism in each year, and atomic and nuclear physics from year 6 on (Ministry of Education, 2014). The previous New Zealand science curriculum mentioned explicitly what children need to know about prisms by the end of year 8 (Minisitry of Education, 1993), but this mention disappeared in the new version (Ministry of Education, 2014). Curie's peak in Adults is interesting. It would be interesting to know how media coverage and mentions in formal education have changed throughout the years about Curie.

5.3.2 Fluency about scientists

Fluency about scientists was only measured once, after the visit. In this case the goal was not to measure any influence from the science centre, but to explore the application of scientific fluency as instrument. When analysing exclusively fluency about scientists, the rescaled mean was M=3.31 (SD=1.03, CI=0.10). This fluency is correlated with age[120] (r_s =.551, p=.01).

While fluency did not depend on gender (independent samples t-test, t(381)=0.544, p=.586, d=0.028, d_{CI}=0.104), familiarity with individual scientists showed remarkable peculiarities. Familiarity with Einstein (Mann-Whitney test, U=17193, p=0.603), Rutherford (U=16085, p=0.117), Newton (U=17367, p=0.738), and da Vinci (U=16806, p=0.371) was

[120] Spearman's correlation was considered more adequate than Pearson's correlation since it was theorized that fluency would increase monotonically with age, not linearly.

not correlated with gender, but female participants (n_F=227) reported more familiarity with Curie (U=15059, p=0.010) than males (n_M=156, Curie: Mdn_M=2, IQR_M=3, CIM=1, Mdn_F=3, IQR_F=3, CIF=0; other scientists: Mdn_M=4, IQR_M=3, CI_M=0, Mdn_F=4, IQR_F=1, CI_F=0)[121].

Overseas visitors (n=32, M_{age}=25.6 years) outscored New Zealanders (n=351, M_{age}=24.7 years) significantly in self-reported familiarity with Einstein (U=4226, p=0.012) and da Vinci (U=4331, p=0.023), but not for Newton (U=4223, p=0.15) or Curie (U=4807, p=0.163). New Zealanders were slightly more familiar with Rutherford (Rutherford: Mdn_{NZ}=3, IQR_{NZ}=2, CI_{NZ}=0, Mdn_{OV}=2, IQR_{OV}=2.25, CI_{OV}=0; other scientists: Mdn_{NZ}=4, IQR_{NZ}=2, CI_{NZ}=0, Mdn_{OV}=4, IQR_{OV}=1, CI_{OV}=0), although not significantly more familiar than overseas visitors (U=4557, p=0.069). It seems likely that a study with a larger sample of overseas visitors would pick up a difference between overseas and domestic visitors in familiarity with this famous New Zealand scientist.

The explanation for overseas visitors outscoring New Zealanders in most scientists may stem from two facts. First, to travel overseas, a visitor's socioeconomic status is likely to be medium to high, and higher socioeconomic status considerably impacts academic achievement, in favour of higher science scores (OECD, 2007).

However, the most important indicator is motivation. Attitudinal and motivational constructs influence science learning (Koballa & Glynn, 2010). Scientific literacy "is regulated by the individual's appreciation, interest, values, and action relative to scientific matters" (OECD, 2006, p. 23). While 26-32% of Otago Museum's visitors come from overseas, only 8-9% of the science centre's overall visitorship comes from overseas (Section 1.3). The fact that some overseas visitors chose a science centre as a venue where they would spend their limited overseas vacation time shows that they were a more motivated population than local visitors. Falk, et al. (2016) administered surveys to more than six thousands adults in 13 countries, finding that science centre goes had significantly higher interest, curiosity and understanding of science than those that do not use science centres.

5.3.3 Fluency in scientific concepts

Fluency in scientific concepts was measured before (M=2.86, SD=0.96, CI=0.10) and after the visit (M=3.30, SD=0.99, CI=0.10). There was no difference between male and female visitors in their fluency in scientific concepts, either before (independent samples t-test, t(381)=1.47, p=.142, d=0.075, d_{CI}=.104) or after the visit (t(381)=1.22, p=.223, d=0.062,

[121] The median for "other scientists" is obtained as a grand median of all four scientists pooled. Notice that it is possible to report a mean in this case, but not for a single item; this is why medians are reported in both cases. Notice also that, unlike in means, CI in a median can be zero.

d_{CI}=.104), but visitors in general increased significantly their fluency in scientific concepts (paired t-test, t(385)=15.56, p<.001, d=0.792, d_{CI}=0.075).

Correlations of fluency in Scientific Concepts and age were calculated with LOESS regressions (Figure 5-3). Notice how Children (8 to 18) present a high positive slope, then between 19 and 25 years there is a local maximum as fluency slightly decreases. Hereafter, it increases monotonically with a gentle slope. Also note this graph's similarities with the LOESS regression of scientific knowledge discussed in Chapter 4 (see Figure 4-3). The explanation is similar: the steeper slope in Children and Adolescents is related to learning during by formal education.

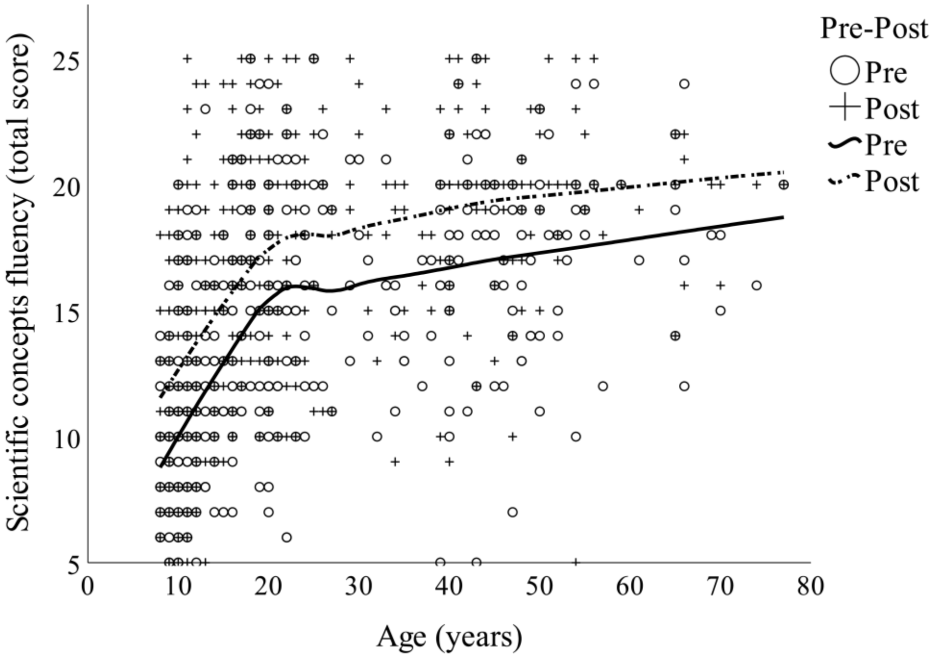


Figure 5-3. Scatter plot with LOESS regressions (α=.70) for fluency in scientific concepts before and after a visit (N=386) to Tūhura.

More research is needed about scientific fluency to confirm if, as these results suggest, it can be a good reflection of scientific knowledge. The importance of this relation is greater than it seems at first sight. "People engaging in everyday learning may not be aware that they are learning" (Fenichel & Schweingruber, 2010, p. 3). But if the correlation is confirmed, it would imply the validation of the assumption of self-reporting being a good measure of actual learning.

The smooth and sustained increase in visitors over 26 years old would indicate that people never stop learning. "The goal of science education in the twenty-first century must be to fully support life-long science learning. Achieving this goal will require attending to the personal learning needs of every citizen across their entire lifespan" (Falk, Storksdieck, et al., 2007, p. 465).

The bump of 19-25 is less clear. It may be that the shift from university, where education is specialized and structured, to the subsequent beginning of working life causes a period of adjustments that directly impact the perception of how much one knows. More research is needed to draw any conclusions.

A LOESS regression doesn't find local linear correlations, but visually helps in detecting where they can be found. The highest local linear correlations happened with Children and Adolescents (8 to 18 years old). The slopes can be linearly approximated as $y_1=1.04 + 0.88x_1$ ($r^2=.350$) before the visit and $y_2=4.56 + 0.78x_2$ ($r^2=.223$) afterwards. y represents the fluency score, x the age, 1 pre-visit and 2 post-visit. Taking the liberty of assuming both slopes are equal to their average (0.83), equating them gives:

$$y_2 = y_1 \rightarrow 4.56 + 0.83x_2 = 1.04 + 0.83x_2$$

which leads to

$$\Delta x = x_2 - x_1 \simeq \frac{4.56 - 1.04}{0.83} \simeq 4\ years = \Delta age$$

In other words, younger visitors in this study left the science centre with a fluency in scientific concepts equivalent to the prior fluency of visitors 4 years older. According to J. Cohen (1988), the effect sizes of both pre and post linear approximations is large (see Table 3-13). However, this result doesn't directly translate into a single visit to a science centre being equivalent to four years of formal education at all. First of all, fluency measured a very small part of the possible scientific knowledge. Moreover, other factors could have a large influence in the outcome, such as the cueing issue and the possibly transitory nature of this increase. In other words, a longitudinal study with not-cued respondents would be needed to know how much of this change is due to the visit itself and how much of it stay after a long period of time. See also 4.2.2 for a lengthier discussion on a similar analysis.

5.4 Self-concept in Science: I'm good at this

Self-concept in science is an individual's general perception of their own abilities in science (see Table 1-3). It is not only an important component of self-beliefs in science

(OECD, 2009; Wilkins, 2000, 2004), but it stands apart from self-beliefs in science by influencing career aspirations in science (Nagengast et al., 2011).

Self-concept in science is similar to self-efficacy in the sense that it is influenced by previous performances (Jansen et al., 2015) and it reinforces with achievement (DeBacker & Nelson, 2000; Huang, 2011; Jansen et al., 2015; Jansen et al., 2014; Marsh & Martin, 2011; Marsh et al., 2012; Wender, 2004; Wilkins, 2004). But it also differentiates itself because the self-assessment is general, about doing well in a given domain (Bong & Skaalvik, 2003; Huang, 2011; Jansen et al., 2015; OECD, 2018; Wilkins, 2004), rather than being related to specific tasks as with self-efficacy. It is also heavily influenced by social comparison (Bong & Skaalvik, 2003; Jansen et al., 2014).

5.4.1 Self-concept: exploratory single items

Self-concept in science was assessed with two exploratory items in Discovery World: a three-point (plus IDK) Likert-type item, SC_{DW1} 'I could explain some science examples to my friends', and a five-point (no IDK) Visual Discrete-type item, SC_{DW2} 'Click on the bunny that best represents yourself in the stair of science understanding'. Both increased only slightly, but significantly after the visit (SC_{DW1}: N_1=224, Mdn_{Pre1}=3, IQR_{Pre1}=4, CI_{Pre1}=1, Mdn_{Post1}=3, IQR_{Post1}=3, CI_{Post1}=1, Wilcoxon-Pratt test, Z_1=3.28, p_1<.001, 38 discordant pairs of 181, r_1=.244; SC_{DW2}: N_2=224, Mdn_{Pre2}=3, IQR_{Pre2}=4, CI_{Pre2}=1, Mdn_{Post2}=3, IQR_{Post2}=3, CI_{Post2}=1, Z_2=3.59, p_2<.001, 39 discordant pairs of 224, r_2=.240).

A five-point (plus IDK) Visual Discrete-type item, 'Click on the penguin that best represents yourself in the stair of science understanding', was used in surveys in Tūhura for comparative purposes. The change was not significant (N=251, Mdn_{Pre}=3, IQR_{Pre}=4, CI_{Pre}=1, Mdn_{Post}=3, IQR_{Post}=4, CI_{Post}=1; Z=1.812, p=.082, 58 discordant pairs of 251, r=.114).

5.4.2 Likert scale and Visual Discrete Scale: a small, but significant difference

The Likert-type scale is a popular self-reporting method, but it is not problem-free. For example, verbal anchors make respondents not perceive the distance between points as equidistant (Lantz, 2013). This poses a dilemma for the validity of testing these scales parametrically (see Chapter 3 and Appendix E).

As an alternative, a Visual Discrete format was developed (see Chapter 3). The stability of self-concept (Lee, 1998) makes it the best among the studied self-beliefs to compare the Likert scale with the Visual Discrete Scale. Based on the exploratory items, a full 7-item scale was created for Tūhura in both formats. Both had five ordinal points and one nominal I Don't Know (IDK) option. Results are very similar in all descriptives (Table 5-1). However, the Visual Discrete Scale detected a small increase in self-concept in science (Figure 5-4). This was significant with a medium effect size (Table 5-1). Therefore, the Visual Discrete Scale is potentially more sensitive to small changes.

Table 5-1
Self-concept in science measured with a Likert scale (LS, N=446) and a Visual Discrete Scale (VDS, N=375) before (pre) and after (post) the visit

		M	**SD**	**CI**	**α**	**%Eli**	**Skew**	**Kurt**	**t-test**	**d**
LS	Pre	3.69	0.81	0.08	.890	4	-0.520	0.165	t(445)=1.55,	d=0.073,
	Post	3.73	0.81	0.08	.919		-0.566	0.298	p=.123	d_{CI}=0.067
VDS	Pre	3.46	0.84	0.08	.898	3	-0.346	-0.129	t(374)=8.33,	d=0.430,
	Post	3.66	0.79	0.08	.913		-0.468	0.031	p<.001	d_{CI}=0.074

SD stands for Standard Deviation, CI stands for Confidence Interval at 95%, %Eli for percentage of eliminated responses due to an excess of missing values (N is the sample size after this deletion), Skew for skewness, Kurt for kurtosis, α for Cronbach's alpha, d for Cohen's d, d_{CI} for confidence interval of d.

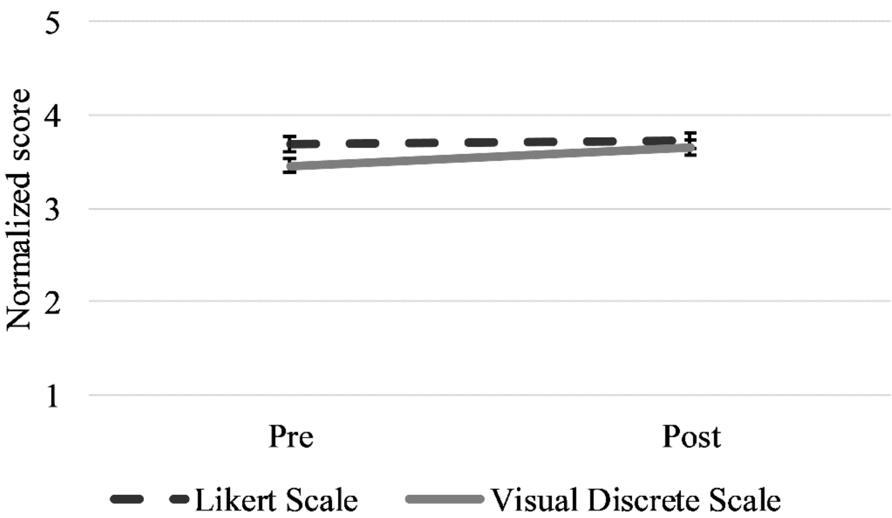


Figure 5-4. Rescaled self-concept in science means before (pre) and after (post) the visit ($N_{\text{Likert scale}}$=446, $N_{\text{Visual Discrete Scale}}$=375).

An anticipated limitation of the Visual Discrete Scale was that it could end up being more difficult for respondents to understand. This would have been reflected by a greater number of missing values. "Missing values seem to occur because the scales are complex, not self-explanatory, and patients are unfamiliar with these tools" (Phan et al., 2012, p. 506). But as can be seen from Table 5-1, Likert scale and Visual Discrete Scale produced similar percentages of missing values.

Regarding Self-Concept in Science, both scales disagreed with Şentürk and Özdemir (2014), who found a relatively large increase in self-concept after visiting a science centre, from a rescaled value of 4.25 to 4.85. However, Kind et al. (2007) used the same instrument in a different setting and self-concept *decreased* significantly from 3.41 to 3.24. Instead, the results of this research agree with Bong and Skaalvik (2003) and Lee (1998) in that self-concept is a fairly stable construct.

Self-concept in science is not correlated with age either before (LS: N=441, r_s=.016, p=.745; Visual Discrete Scale: N=380, r_s=-.049, p=.346) or after the visit (LS: N=441, r_s=-.080, p=.094; Visual Discrete Scale: N=380, r_s=-.013, p=.804).

5.4.3 Self-concept: back to a single item

As was explained above and in Chapter 3, the study of self-concept started out with two single exploratory items: one in Likert format, and one in Visual Discrete format. The reason for using single items was mainly related to the urgent need to collect data from Discovery World before it closed. However, the items were not arbitrarily created; they were the product of a careful and thorough discussion by a panel of experts (see Chapter 3). This section is dedicated to see if self-concept in science can actually be measured with a single item. Self-concept in science was measured with a Likert scale and a Visual Discrete Scale. A factor analysis confirmed that both scales were unidimensional (Table 5-2).

Table 5-2
Factor analysis by principal components of self-concept in science as measured with Likert scale (LS, N=446) and Visual Discrete Scale (VDS, N=375)

		Bartlett's test	**KMO**	**# of components**	**Variance explained**
LS	Pre	$\chi^2(446,21)=1764$, p<.001	.911	1	63%
	Post	$\chi^2(446,21)=2344$, p<.001	.925	1	70%
VDS	Pre	$\chi^2(375,21)=1461$, p<.001	.926	1	63%
	Post	$\chi^2(375,21)=1726$, p<.001	.930	1	67%

KMO stands for Kaiser-Meyer Olkin.

The component matrices (Table 5-3) contain estimates of the correlations between each of the variables (items) and the estimated component (self-concept). Most items consistently ranked high, indicating that several of them could be used to assess self-concept on their own.

Table 5-3
Component matrices for self-concept in science measured with Likert scale (N=446) and Visual Discrete Scale (N=375)

	Likert scale		**Visual Discrete Scale**	
	Pre	Post	Pre	Post
Item 1 - Understanding of science	.860	.902	.851	.871
Item 2 - Explain science examples	.786	.832	.798	.815
Item 3 - Learn science fast	.841	.887	.823	.845
Item 4 - Solve math problems	.498	.594	.570	.612
Item 5 - Solve science problems w/o math	.790	.833	.859	.869
Item 6 - Understand new science ideas	.822	.851	.796	.833
Item 7 - Do well in science	.873	.897	.839	.870

Specific wording of each item depending on the format can be found in Table 3-2.

Item 1, about the 'level of understanding of science' and Item 7, about 'doing well in science', are the best for use as single items. Notice that Item 1[122] was one the items approved by the panel of experts, which gives a vote of confidence to both the results obtained from Discovery World and the experts working in the research. Item 7 is the original single item used to measure self-concept in science by the International Association for the Evaluation of Educational Achievement (IEA), in the first two Trends in International Mathematics and Science Study (IEA, 1994, 1998). This item was compared to a later-developed full scale (IEA, 2002) by Wilkins (2004) who determined thesingle item to be a valid measure of the whole scale.

Diamantopoulos, Sarstedt, Fuchs, Wilczynski, and Kaiser (2012) state that single items are a reasonably safe option when the sample size is smaller than 50, or when inter-item correlations are above .80 or Cronbach's alpha is higher than .90. Both items 1 and 7 are above the inter-item correlation cut-off in both scales (Table 5-3) and alpha is very close to .90 in both scales (Table 5-1).

5.5 Why the Likert Scale Didn't Detect a Change

5.5.1 The one-size-fits-all behaviour

There are phenomena in the scales that cannot be seen from the summarized results and deserve closer attention. 'rescaled score 4 responses' are defined as to those responses for which rescaled score was 4 and 'monotone-agree responses' to a subset of rescaled score

[122] The other is Item 2, which also got high levels of correlation, but doesn't go over .80 in all cases, as recommended by Diamantopoulos, Sarstedt, Fuchs, Wilczynski, and Kaiser (2012).

4 where all the answers in a given scale were ‘Agree’ (option 4). Figure 5-5 shows the difference from Pre (before the visit) to Post (after the visit) rescaled scores of the three self-beliefs in science measured in this thesis. The difference is presented as percentages, centred around rescaled score 4.

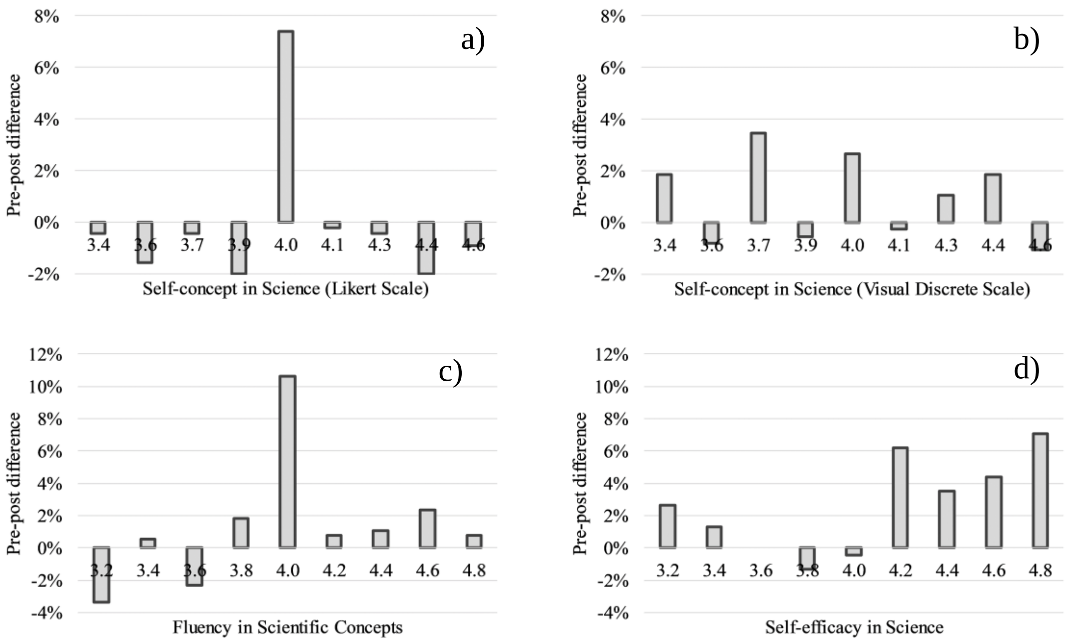


Figure 5-5. Percentage score differences from before (pre) to after (post) a visit to Tūhura around the rescaled score 4 in: a) self-concept in science (Likert scale, N=446), b) self-concept in science (Visual Discrete Scale, N=375), c) scientific fluency (N=386), d) self-efficacy in science (N=227).

Monotone agree responses for self-concept in Likert scale were 23 (5%) before the visit. rescaled score 4 were 38 (9%). After the visit, these figures increased to 59 (13%) monotone agree and 71 (16%) rescaled score 4. Before the visit monotone agree responses represented half of all rescaled score 4 responses, but after the visit Monotone Responses were more than three quarters. This considerable relative increment of monotone agree responses is called, by the researcher, the *one-size-fits-all behaviour*.

When self-concept was measured with the Visual Discrete Scale, monotone agree passed from 14 (4%) to 29 (8%) and rescaled score 4 from 28 (7%) to 38 (10%). The one-size-fits-all behaviour is smaller.

In fluency in scientific concepts, rescaled score 4 responses increased from 21 (6%) to 62 (16%). The change in monotone agree responses was from 15 visitors (4%) to 53 (14%). Once again, a significant one-size-fits-all behaviour.

Self-efficacy in science is the only self-belief where no one-size-fits-all behaviour was detected whatsoever. Rescaled score 4 responses even decreased from 11 (5%) to 10 (4%) (monotone agree from 4 (2%) to 1 (0%)). The reason might be that self-efficacy pushes respondents to the extremes (Section 5.2). In that case, the one-size-fits-all behaviour would happen in NS5 responses, which indeed increased from 4 (2%) to 26 (12%). Notice that monotone agree coincides with NS5 in this case. However, this is an unlikely explanation of the shift, because the origin of the peak wouldn't correspond to remembering the items and finding one response that fits all of them. It is more likely that the nature of self-efficacy avoids the general behaviour of responding with the same answer, as each item feels more 'individual' than with self-concept, where all items are similar in nature.

5.5.2 The sponge effect

In the previous section, it was noticed that rescaled score 4 responses can sometimes noticeably increase after the visit, partly because of monotone agree responses. However, notice how self-concept in science, as measured by Visual Discrete Scale (Figure 5-5 b), presents a second small peak in the rescaled score 3.7. This score cannot be obtained from monotone responses. Thus, an increase in score may be due to something beyond a simple lack of interest in answering (i.e., choosing the same answer just to finish quick).

One interesting characteristic of self-concept in science, when measured with the Likert scale (Figure 5-5 a), is that the monotone agree responses increased while the scores lowered around it[123]. This pulling from its neighbours is named, by the researcher, the *Sponge effect.* It is the combination of one-size-fits-all behaviour and Sponge effect which stopped the Likert scale from measuring changes in a fairly stable construct.

5.5.3 Explaining the one-size-fits-all behaviour and the sponge effect

This behaviour of answering with a range of options before the visit, and then answering with a single option for each item in the whole scale after the visit, was not found in the literature. The closest examples were a short mention of choosing a particular column in tests of academic mastery as a peculiar bias (J. W. Osborne & Blanchard, 2011), and a case where a high percentage of children chose 5 out 5 in both pre and post ratings (Hall et al., 2016).

Several factors may contribute to the above phenomenon. First, it is necessary to acknowledge that, according to the 'classical test theory', every measurement (observed

[123] It is only shown from 3.4 to 4.6, but it actually happened from rescaled scores 3.3 to 5.0

score) for a person is an additive composite of a true score specific to the person plus an error score. Neither the person's true score, nor the error score are directly observable; measurements can differ under different conditions (Brennan & Lee, 2018). In addition, after experiencing potentially overwhelming plethora of information (which might happen in Tūhura), respondents tend to issue a quick, self-reported response of their personality to get on with the task (Paulhus & Vazire, 2007). Last, but not least, the 'testing effect' postulates that respondents to multiple-choice questions generally improve performance on a later test due to memory (Kromann, Bohnstedt, Jensen, & Ringsted, 2010; Roediger & Marsh, 2005). It is also important to consider that respondent's confidence is based on the ease with which potential answers come to mind (Kelley & Lindsay, 1993).

Some smart respondents may have realized that the questions in the scale were closely related, like asking something in common. When filling out the second part, they could have remembered that and, to avoid more mental exertion after the visit, they answered them with the one answer that best represented their overall thinking. For instance, 3,4,4,4,4,3,4 and 4,4,4,4,5,4,4 before the visit would both become 4,4,4,4,4,4,4 after the visit. They weren't trying to avoid giving thoughtful answers, only to minimize the cognitive workload. Notice how this reasoning would explain not only the one-size-fits-all behaviour, but the Sponge effect. To sustain that Monotone Answer respondents gave honest feedback, open questions in their surveys were analysed individually and no skipping behaviour was found.

A clue to why these phenomena happened in the Likert scale, but not in the Visual Discrete Scale, comes with the following quote: "Repetitiveness of the items on a questionnaire may decrease a respondent's motivation to maintain the cognitive effort required to provide optimal answers and increase the desire to satisfice by responding in a nondifferentiated manner or stylistically" (MacKenzie & Podsakoff, 2012, p. 551). This quote is related to the fact that using the same scale format for all constructs in a questionnaire can produce method bias (Podsakoff et al., 2003), something that didn't happen here because the surveys were very short. But, it could be that this repetitiveness goes beyond an individual questionnaire and people have become habituated to Likert-type scales. If so, they find the format familiar, which produces a confidence that all the questions can be answered with a single option. Since the Visual Discrete Scale is new, they may have lacked confidence from unfamiliarity, and felt more inclined to read all the questions again.

Also, cute products make consumers more indulgent in consumption choices (M. Scott & Nenkov, 2013). In this case, visitors didn't spend money, but the cuteness of the penguin in the Visual Discrete Scale could have made them more indulgent in time spent, which translated into giving more thoughtful responses.

If the above explanations are correct, it would imply a couple of things. First, the phenomenon was not reported before in the literature because in order to detect them, the construct being measured needs to be extremely stable and be measured before and after an intervention. It has already been shown in Chapter 1 how rarely these measures are done. It doesn't mean that the effect only happens when the construct is stable. It may be present in any other constructs (including scientific fluency and self-efficacy in science), but any relatively large change in scores from pre to post would hide it. Second, Visual Discrete Scale may be of use not only as an alternative to Likert-type scales, but as a companion to Likert-type scales in the same questionnaire. Variation of scale properties is an effective remedy for controlling method bias (Podsakoff et al., 2012).

5.6 Gender Differences in Self-beliefs in Science

The three self-beliefs before the visit and their respective changes after the visit were tested for gender differences (Table 5-4). The only self-belief that didn't show gender differences before the visit is scientific fluency. Why there is no gender gap in this case, but there is in other self-beliefs and in scientific knowledge (see Chapter 4), requires further research to be understood.

In agreement with Mostafa (2019), a gender gap is clearly identified in self-efficacy in science. A study by the OECD (2014) didn't find a gender gap in self-concept in PISA in New Zealand. However, in this research, male visitors were found to present a higher self-concept in science than females, which agrees with Jansen et al. (2015) and Kurtz-Costes et al. (2008).

Table 5-4
Gender differences in self-beliefs in science

	Self-concept in science (Likert scale)					
	Males (n=179)			**Females (n=263)**		
Pre	M=26.9,	SD=5.7,		M=25.1,	SD=5.5,	
	CI=0.8		t(178)=0.147, p=.883,	CI=0.7		t(262)=2.09, p=.038,
Post	M=26.9,	SD=5.8,	d=0.000, d_{CI}=0.106	M=25.5,	SD=5.6,	d=0.072, d_{CI}=0.087
	CI=0.9			CI=0.7		
	Gender difference before the visit: t(440)=3.34, p<.001, d=0.159, d_{CI}=0.097					
	Self-concept in science (Visual Discrete Scale)					
	Males (n=153)			**Females (n=219)**		
Pre	M=25.4,	SD=5.2,		M=23.4,	SD=6.1,	
	CI=0.8		t(152)=5.66, p<.001,	CI=0.8		t(218)=6.33, p<.001,
Post	M=26.7,	SD=5.4,	d=0.245, d_{CI}=0.115	M=24.9,	SD=5.5,	d=0.258, d_{CI}=0.096
	CI=0.9			CI=0.7		
	Gender difference before the visit [a]: t(355)=3.49, p=.001, d=0.363, d_{CI}=0.104					
	Fluency in scientific concepts					
	Males (n=156)			**Females (n=227)**		
Pre	M=14.7,	SD=4.9,		M=14.0,	SD=4.7,	
	CI=0.8		t(155)=8.53, p<.001,	CI=0.6		t(226)=13.4, p<.001,
Post	M=16.9,	SD=5.1,	d=0.440, d_{CI}=0.115	M=16.2,	SD=4.8,	d=0.463, d_{CI}=0.095
	CI=0.8			CI=0.6		
	Gender difference before the visit: t(381)=1.47, p=.142, d=0.146, d_{CI}=0.104					
	Self-efficacy in science					
	Males (n=85)			**Females (n=137)**		
Pre	M=14.5,	SD=6.0,		M=11.7,	SD=5.1,	
	CI=1.3		t(84)=10.8, p<.001,	CI=0.9		t(126)=14.2, p<.001,
Post	M=19.5,	SD=5.3,	d=0.883, d_{CI}=0.161	M=16.8,	SD=5.6,	d=0.952, d_{CI}=0.127
	CI=1.1			CI=0.9		
	Gender difference before the visit [a]: t(158)=3.57, p<.001, d=0.503, d_{CI}=0.140					

[a] These samples didn't pass the Levene's test for equality of variances (p-value was less than .05) and the statistic reported correspond to the assumption of equal variances not assumed.

Besides confirming the existence of a gender gap, it is interesting to note that the changes produced by visiting the science centre tend to be similar in both genders, as Şentürk and Özdemir (2014) already noticed in self-concept.

An interesting question is when the prior gaps start. The sample size of self-efficacy in science is not big enough to split data into age groups, but the sample size of self-concept in science (Likert scale)[124] is. Notice in Table 5-5 that Children do not show a gender gap, while one is clear in both Young Adults and Mature Adults. Unfortunately, the sample size of male Adolescents (n=20) is too small to conduct valid statistical tests (see Chapter 3). Why PISA didn't detect a gender gap in self-concept in science in New Zealand will be further discussed in the next chapter.

Table 5-5
Gender differences per cohort in self-concept in science (Likert scale) before the visit to Tūhura (N=442)

		Descriptives	**Mann-Whitney U test**
Children (8-12)	Males (n=51) Females (n=52)	Mdn=26 Mdn=26	U=1291, p=.819, r=.023
Adolescents (13-18)	Males (n=20) Females (n=55)	Mdn=27.5 Mdn=26	NA [a]
Young Adults (19-40)	Males (n=57) Females (n=94)	Mdn=28 Mdn=25.5	U=1909, p=.003, r=.241
Mature Adults (41+)	Males (n=50) Females (n=58)	Mdn=28.5 Mdn=25.5	U=1010, p=.006, r=.261

[a] The sample size of adolescent males is too small to conduct a valid test (see Chapter 3).

Lastly, the percentage of the total number of IDK (I Don't Know) selections and the number of respondents that selected at least one IDK in any given scale[125] was expected to be an indirect measure of confidence. However, the only scale that showed a significant gender difference (males selecting more IDK) was self-efficacy in science. But it is speculated that many of the IDK responses in this scale actually correspond to respondents not knowing anything about the concept asked, and instead of selecting "Not confident at all", they went for IDK.

5.7 Towards the Validation of Self-reporting to Assess Knowledge

5.7.1 Fluency and self-assessment of learning

In Chapter 4, the results of directly asking visitors whether they learnt something new were discussed. Although this question was open to any field, fluency in scientific concepts

[124] The advantages of the Visual Discrete Scale over the Likert scale are to detect changes after the visit. Before the visit, both are equivalent, and Likert scale had a larger sample size.
[125] After the cleaning, but before the imputation and with no listwise deletion. IDK in scientific knowledge is not considered in relation to confidence, but of lack of knowledge and is not analysed in terms of IDK.

was narrowed for this research to the topics of light and electromagnetism. Still, it is interesting to compare them.

Those who believed they learnt something new (n=276) significantly increased (t=16.62, df=275, p<.001, d=1.000, d_{CI}=0.090) their fluency in scientific concepts mean from M=2.91 (SD=0.91, CI=0.11) to M=3.4 (SD=0.89, CI=0.11). Those that believed they didn't learn anything new (n=78) also increased this mean significantly (t(77)=2.48, p=.015, d=0.281, d_{CI}=0.161), but only from M=2.87 (SD=1.09, CI=0.24) to M=3.03 (SD=1.19, CI=0.26). While both changes are significant, the increase in fluency and the effect size are considerably larger in those who believed they learnt something.

To confirm the difference in fluency in scientific concepts between the group of those that reported learning (n=276) and those that didn't (n=78), an independent samples t-test[126] was conducted on both groups. The groups were not statistically different before the visit (t(109)=0.271, p=.787, d=0.040, d_{CI}=0.128), but they were afterwards (t(103)=2.654, p=.009, d=0.352, d_{CI}=0.129).

5.7.2 All roads lead to learning

In Chapter 1, the multiple possible issues with self-reporting that can lead to inaccuracy were discussed. Flawed self-assessment can come from underestimation (Parkman, 2016; Sherman, 2013), overestimation (Dunning, 2011; Dunning et al., 2004; Mahmood, 2016; Schlösser et al., 2013), or just incorrect recollection (Kelley & Lindsay, 1993; Ladwig et al., 2012; Mbewe et al., 2010; W.-C. Wang et al., 2016).

This research assessed scientific knowledge objectively (see Chapter 4), but also measured self-beliefs in science. Figure 5-6 shows the mean scores of scientific knowledge, self-efficacy in science, and fluency in scientific concepts before and after the visit. The three instruments come from different surveys, but demographics were fairly similar, and all instruments were related to light and electromagnetism (see Chapter 3).

[126] Both cases were evaluated with equity of variances not assumed.

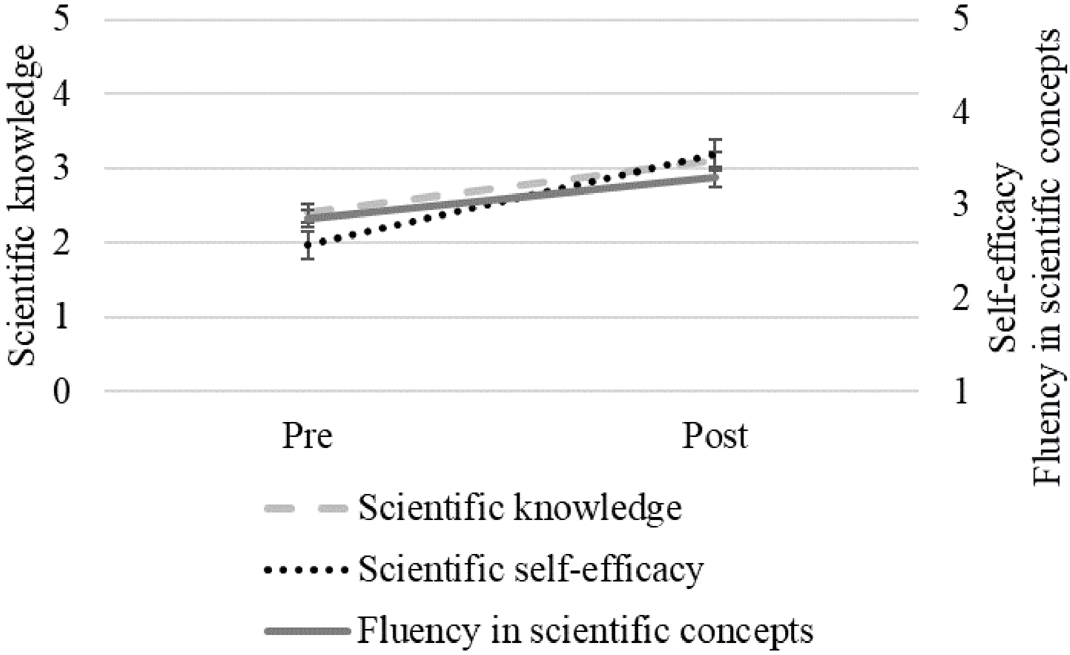


Figure 5-6. Scores of scientific knowledge (N=456), self-efficacy in science (N=227) and fluency in scientific concepts (N=386) before (pre) and after (post) visiting the science centre. Error bars represent confidence intervals at 95%.

The three of them behave similarly. Self-efficacy in science has a stronger slope, falling a bit off from what would be expected in order to resemble scientific knowledge. This is due to the previously explained effect of respondents dichotomizing into *can't* before the visit and *can* after the visit.

Fluency in scientific concepts turned out to be a better reflection of scientific knowledge, due to the degree that their confidence intervals overlap before and after. Since these two instruments were conducted with different samples and the items were not paired (as it was done with Likert scale and Visual Discrete Scale), further research is needed to determine if fluency in scientific concepts can truly reflect scientific knowledge. However, these are promising results in that direction.

5.8 Conclusions

Three self-beliefs in science were tested in Tūhura. Self-concept in science, as expected, was so stable that it barely changed from before to after the visit. Nonetheless, this characteristic worked perfectly to test an alternative to Likert-type scales. The newly-developed Visual Discrete Scale was able to detect the small change in this construct. More research is needed around the Visual Discrete Scale, but the results here indicate that it is potentially more sensitive than Likert-type scales in detecting small changes.

Self-efficacy in science, as expected, was heavily influenced by the visit. What was not expected was that many visitors would dichotomize items, changing from *cannot* (lower end of the scale) before the visit, to *can* (upper end of the scale) after the visit. The instrument requires more work to avoid this behaviour.

The third self-belief, scientific fluency, is new to the literature, but it already showed interesting results. The instrument was designed to measure the individual's perceived knowledge. Comparison of the results with those of the scientific knowledge instrument indicates that scientific fluency may be a reliable way to measure scientific knowledge through self-report, perhaps even better than self-efficacy.

In terms of the research questions, self-beliefs in science can be reliably measured with the instruments developed in this research. Self-concept is so stable that it required a novel format to detect minimal changes; the Visual Discrete Scale rose to that challenge (RQ1). Only malleable self-beliefs, such as self-efficacy and fluency, are clearly influenced by visiting a science centre (RQ2). Self-concept seems to also increase due to the visit, but given the small change, more research is needed to confirm this.

Regarding RQ3, no significant differences were found age-wise in self-concept in science and self-efficacy in science, but fluency in scientific concepts increases rapidly during the school age and later increases at a slower pace. Gender-wise, things were the other way around. While fluency didn't show any significant difference, ratings of self-concept and self-efficacy were lower in females than males. However, when analysing self-concept by cohorts, no gap was discovered between girls and boys in the youngest group participants.The next chapter will present evidence of science engagement at the science centre and discuss the gender gap from a perspective based on engagement.

Chapter 1:
INFORMAL LEARNING OF SCIENCE

Chapter 2:
OTAGO MUSEUM SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE: Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE: To know that you know

Chapter 6:
SCIENCE ENGAGEMENT: The joy of learning

Chapter 7:
EXHIBITS AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 6: SCIENCE ENGAGEMENT: The joy of learning

6.1 Introduction

In previous chapters, scientific knowledge and self-concept in science were discussed. This chapter examines the third aspect of scientific literacy: science engagement. Engagement is a complex and highly personal matter where qualitative approaches can be particularly useful to examine its broad spectrum.

The first section is all about how visitors become engaged with science at this study's science centres. The ultimate importance of engagement is clear in this note left by a visitor on the Questions and Answers board. "I am eleven. At how old can I be a scientist?" The second section addresses a fundamental goal of the science centre redevelopment: making Tūhura a place for all generations. Tūhura's capacity to engage children and adults alike is reviewed.The last section discusses the gender gaps in scientific knowledge and self-concept in science that were identified in the previous two chapters. Here, they are not analysed with respect to gender, but how they stem from personal choice based on study options and engagement with those options.

6.2 Engaging with Science

Science centres have been criticized for promoting fun and enjoyment rather than education and science (Fors, 2006). However, visiting a science centre is usually a freely-made choice, motivated by parents' or one's own interests (Falk & Needham, 2013). Engagement in settings such as these becomes a key element in informal learning (Ainley & Ainley, 2011; Boekaerts, 2010; Braund, 2012; Falk & Dierking, 2016; Lin et al., 2012; Sefton-Green, 2012).

6.2.1 An engaging stay

<u>Visit time</u>

Visitors stay at an exhibition as long as they remain engaged (Serrell, 1997). Visit time can be used as a measure of the value the individual assigns to the visit (Jacobsen, 2016). Therefore, it is a proxy measure of engagement. Table 6-1 presents the visit time before and after the redevelopment. To minimize the influence of seeing the Science Show or the planetarium show, visit times from visitors who did these activities were removed[127]. The amount of time spent in the Tropical Forest can only be acknowledged, not removed[128].

Tūhura can be considered more engaging, as visitors spend significantly more time there than they did at Discovery World, including first-time visitors. Moreover, there were two things that happen at Tūhura that didn't happen at Discovery World. First, the more often visitors come, the longer they stay. Second, visitors who interact with the exhibits stay longer than those who do not interact.

Table 6-1
Visit time (minutes:seconds) at the science centre before (Discovery World) and after the redevelopment (Tūhura)

		Discovery World				**Tūhura**			
		n	**M**	**SD**	**CI**	**n**	**M**	**SD**	**CI**
All		159	57:28	21:00	3:14	744	66:16	25:10	1:48
		*t(264)=4.618, p<.001, d=0.380, d_{CI}=0.088							
First-time visitors[129]	1	83	55:20	18:30	7:54	151	60:28	18:25	2:54
		t(232)=2.040, p=.043, d=1.35, d_{CI}=0.17							
Number of visits[130]	1	83	55:20	18:30	3:57	470	64:29	23:19	2:05
	2-3	35	58:27	24:49	8:12	138	68:51	27:54	4:42
	4+	41	60:58	22:13	6:48	58	75:47	26:33	6:50
		F(2,156)=1.042, p=.355, η^2=.013				*&F(2,131)=5.615, p=.005, η^2=.019			
Interact with exhibits	N	38	61:33	25:01	7:56	52	54:26	24:14	6:34
	Y	121	56:11	19:30	3:28	692	67:09	25:02	1:52
		*t(51.9)=1.209, p=.232, d=0.239, d_{CI}=0.186				t(742)=3.542, p<.001, d=0.516, F=0.144			

*Tested with Welch ANOVA due to homogeneity of variances fail.
&Only the difference from 1 to 4+ visits was significant (Games-Howell test, p=.008).
Number of visits and interaction with exhibits were also taken as variables.

Serrell's sweep rate index

Serrell's sweep rate index (SRI) can easily compare similar venues by dividing the exhibition's floor area (in sq.ft) by the total time spent (in minutes) (Jacobsen, 2016; Randi

[127] Science Show was only performed at Discovery World. The Planetarium shared entrance with the science centre only at Tūhura.
[128] Some museum floor staff comment that, at Discovery World, it was quite common for visitors to spend most of their time in the Tropical Forest. At Tūhura, they started noticing that people stayed for shorter lengths of time, to spend more at the exhibits. If so, the difference in exhibit-exclusive visit times may be larger than the difference in times found here.
[129] In the case of Tūhura visitors, only those that didn't visit Discovery World were considered.
[130] Visits were counted individually 1, 2, 3, 4+ in Tūhura, but combined in 2-3 visits for ease of comparison.

Korn & Associates Inc., 2006; Serrell, 1997, 2010). The smaller the SRI value, the more time visitors are spending in the area.

In Chapter 2 it was mentioned that the exhibits area increased from 393 to 654 sq-m after the redevelopment. However, to calculate Serrell's SRI, we need to consider the entire area visitors spend time in, including the Tropical Forest, Tropical Forest Discovery Zone, and the Welcome area. The total floor area is 710 sq.m (7,642 sq.ft) in Discovery World and 869 sq.m (9,354 sq.ft) in Tūhura. Both science centres were larger than the typical exhibition space (5,000 sq.ft) in most large museums (McKenna-Cress & Kamien, 2013)

SRI changed from 133 in Discovery World to 141 in Tūhura. As a comparison, Serrell (1997) calculated the average SRI in large non-diorama exhibitions (>3,900 sq.ft) as 400. Both science centres have an excellent SRI, much better than the average exhibition.

6.2.2 Self-reported engagement

Interest is the gateway to learning (Falk & Dierking, 2016), and something that can trigger this interest is positive emotions. Feeling the joy of positive emotions encourages to open the mind and try new things creatively to expand the understanding of our world (Harré, 2001). Since engagement is associated with positive emotions, and these emotions with learning, assessing engagement with science at the exhibition is a must.

A 5-point (plus IDK) Likert-type scale was used in Tūhura[131] to assess science engagement. This scale was created based on scales produced by Füchslin et al. (2018), Weinburgh and Steele (2000), Hodzi (1992), and Guzey et al. (2014). The pilot included the scale before and after the visit, but since no change was found (see Chapter 3), it was subsequently only asked before the visit.

The mean of 4.0 (CI=0.1, SD=0.8, N=369) denotes a high level of engagement with no difference in terms of gender (Mann-Whitney U test, U=16088, p=.936, r=0.004, n_M=149, n_F=217) or age (Spearman's rho, r_s=.082, p=.117, N=365). Visit time had a small positive correlation with engagement (r_s=.134, p=.010, n=367), significant at the 0.05 level, but it is likely this figure is affected by having individuals with different levels of engagement in one group of visitors that share the same visit time. No difference was found between males (n=149, M=4.0) and females (n=217, M=4.0) (Mann-Whitney U=16,088, p=.936, r=.004). However, a ceiling effect (see Chapter 3) could be affecting comparability.

[131] Three exploratory 3-point (plus IDK) Likert-type items related to science engagement and appreciation of science were asked in Discovery World. "I think science museums are important", "Scientists' work is important to me" and "I expect to enjoy / enjoyed Discovery World". Likely due to a ceiling effect caused by using only three points, none of these items showed a significant increment and will not be discussed further.

6.2.3 In their own three words

Based on the work by Longnecker et al. (2014), visitors were asked to describe the science centre in three words before and after their visit. Word clouds were created for each science centre before and after (Figure 6-1). Size and shades of grey correspond to the frequency of word use. For improved readability, only words with a frequency of at least 2 are plotted. The sample size (N) of respondents (R) and of words (W) for Discovery World (D) and Tūhura (T) before (B) and after (A) the visit, were as follows: $N_{RDB}=222$, $N_{WDB}=647$, $N_{RDA}=221$, $N_{WDA}=653$, $N_{RTB}=387$, $N_{WTB}=1126$, $N_{RTA}=410$, $N_{WTA}=1197$.

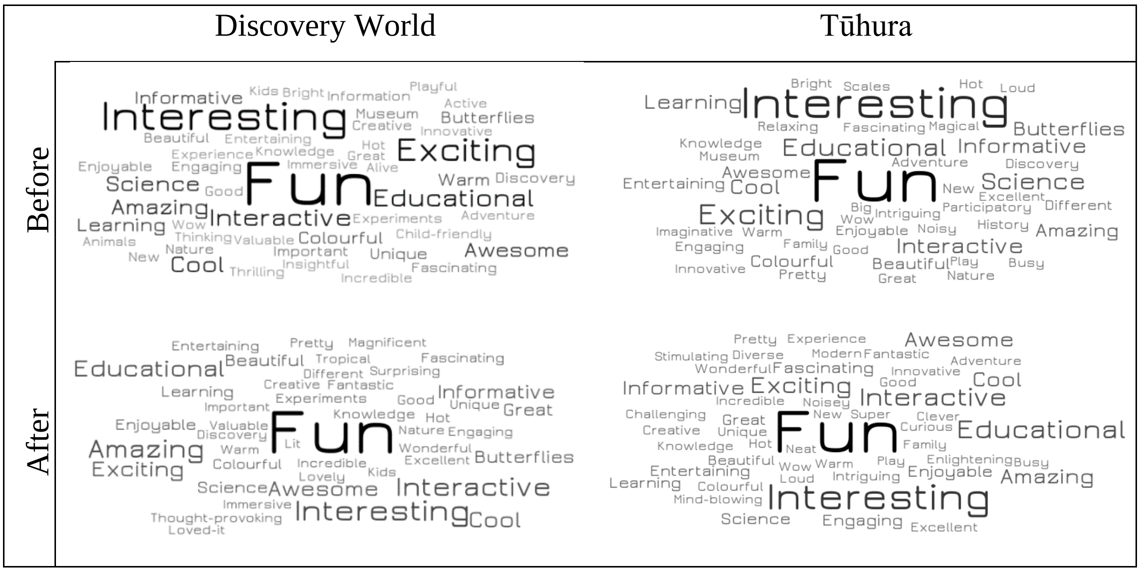


Figure 6-1. Word clouds of the three words visitors chose to describe Discovery World and Tūhura before and after the visit.

Fun and *Interesting* were the words that occupied the first and second places in all cases. *Exciting* and *Educational* were also among the most popular words for both science centres before and after the visit. This coincides with Longnecker et al. (2014), who found that the five most common words to describe a science event were *Fun*, *Interesting*, *Inspiring*, *Educational*, *Exciting*, in that order.

The three words that increased their frequency the most in Discovery World were *Interactive* (28 → 41), *Amazing* (22 → 37) and *Great* (3 → 15). In Tūhura they were *Fun* (215 → 258), *Interactive* (41 → 70) and *Awesome* (20 → 44). The ones that decreased the most in Discovery World were *Interesting* (72 → 46), *Exciting* (49 → 26) and *Science* (22 → 12). In Tūhura they were *Butterflies* (33 → 1), *Science* (50 → 21) and *Learning* (34 → 16). At first sight it may seem like the science centres had a negative impact on science engagement, decreasing the frequency of concepts like *Science* and *Learning*, but a simpler answer is that these words could be linked more to reasoning than to feeling, and words like *Amazing* and *Awesome* are on the opposite side of the spectrum. The interpretation is that visitors come with expectations of a cognitive experience, with science in their brains, but leave with science in a deeper part of the being: emotions. That's what science engagement is about. Emotions are intertwined with science (Koppman, Cain, & Leahey, 2015) and are the link to an effective science communication (Davies et al., 2019; Roeser, 2012).

Children and Adolescents described both science centres after the visit more vividly, with descriptions like *Cool*, *Amazing*, *Awesome,* and *Exciting*, as opposed to adults describing them more rationally, with words like *Interactive*, *Educational,* and *Informative*. This coincides with Randi Korn & Associates Inc. (2006), who found that most children describe science exhibitions as *Cool* and *Fun,* while adults appreciate the opportunity to interact with the family and help the children think creatively.

The relative weight of the exhibits and Tropical Forest in the choice of words is unknown, but the decreasing *Butterflies* in Tūhura indicates the science exhibits outside of the Tropical Forest gained prominence. The reason for the drop in *Exciting* in Discovery World is unclear, but may be related to the exhibits' old age (see Chapter 2).

6.2.4 Appreciation for the science centre

Survey open questions allowed visitors to show both how deep and broad their engagement is. Their comments were categorised, and *Enjoyment and Appreciation* was the most popular one in Discovery World (60%, 24)[132]. For example, "It was lotsss [sic] of fun" (DW, F, 8-12). Discovery World comments on Enjoyment and Appreciation were largely influenced by the Tropical Forest (23%, 9). "It was very fun here and [I] loved seeing the different [butterflies]" (DW, F, 8-12).

There were a lot fewer mentions of the Tropical Forest in Tūhura comments (3%, 7), but the proportion of visitor comments that fit into the Enjoyment and Appreciation category was similar (54%, 118). However, Tūhura visitors, especially adults, tended to express their enjoyment with more intensity than in Discovery World. Some examples are:

[132] The comments were supposed to be about the exhibits, but the percentages of visitors that talk about the Tropical Forest instead of the exhibits are similar to the percentages of visitors that didn't interact with the exhibits at both versions of the science centre (see Chapter 3).

- Children. “it was really fun and a different way of learning. I don´t think there is anything to add” (T, M, 10). “No comments or suggestions that I can say because it has been so fun for me today” (T, F, 11).
- Adolescents. “I think that your name ‘Tūhura’ needs some pizzaz[z], perhaps ‘Tūhura: The most slendiferous pinnacle of scientific wonders to undoubtedly amaze and enthrall through loquacious, gregarious, garrulous designs of incredibly, indubitably largely magnanimous optimism and fun towards scientific adventure! Just a suggestion” (T, M, 13).
- Young Adults. “This place is flippin' awesome! I'll be back, and will definitely continue to recommend this area to everyone!” (T, F, 25). “Much more here than I expected, and stuff I had never thought about (T, M, 33).
- Mature Adults. “We had an amazing time[133] Will definitely be visiting again. Thank you” (T, F, 41). “It is a great interactive space. As an adult I almost found it overstimulating with motion, light, noise. Awesome to visit” (T, F, 50).

The second most common category of comments in both science centres was *Suggestions*. In Discovery World (25%, 10), it was mainly to ask for an upgrade: “Put in some new equipment” (DW, M, 8-12). “Make it more/have more places for teens” (DW, F, 8-12). “Make the descriptions shorter and more fun to read” (DW, F, 13-18). “More science experiment, less game” (DW, M, 19-40). “Be sure to fix the broken exhibits as soon as you are able!” (DW, M, 41+). In Tūhura (27%, 59), fewer suggestions involved changes. “More light around the writing for exhibits” (T, F, 18). Most were about adding things. “Please make an exhibition about earth spin[n]ing” (T, M, 10).

Some suggestions show that the community not only knows what they want, but could contribute to a Think Tank. “Biomechanics = forces and human motion. More of these concepts could be explained with the skeleton on the bike” (T, F, 43). Visitors can be contributors with a museum. It can be a source of engagement for the visitors and generate outcomes different from what staff alone could do (N. Simon, 2010)

> Need to reinstate small SOUNDPROOF [uppercase in original] Space for science interpreters to do science demos with strong focus on specific science themes...not just frozen ice cream and popping hydrogen balloons as per old discovery world... Could use some of the space from the planetarium foyer area. other ideas also [visitor’s phone number removed by the researcher]. (T, M, 59)

But visitors’ comments go beyond suggesting improvements and offering thanks for a great time, they go as deep as the science centre touching lives. “Keep doing what you do.

[133] There were three emojis inserted here but it was not possible to know which ones from raw data.

It’s a valuable tool for dorky parents like me who slept through Science and want to teach their kids, as well as themselves, about the awesome world of science!” (DW, F, 19-40). Even more than individuals’ lives, the science centre can have an impact on the community through visitors. “It was a fantastic experience and resource and gave me ideas to share with colleagues and tamariki[134] back at my kindergarten, and my grandchildren. Thanks” (T, F, 61).

Tūhura visitors who had previously visited Discovery World (n=150) were asked what they missed from Discovery World. Almost half (46%, 69) indicated they did not miss anything. “No, [I like] the new exhibits, seeing the same stuff too many times gets boring so it is nice to have change” (T, M, 17). “From an [adult’s] perspective, this is much better. It is still extremely interactive for all, and I feel like the learning experiences are more beneficial for adults (as well as children) than ones that were offered at [Discovery World]” (T, F, 25). The most missed exhibits were *Foot Stomping Piano* (13%, 20), *MindBall* (7%, 11) and *Air Hockey Table* (7%, 11). Characteristics of engaging exhibits will be discussed in Chapter 7.

6.2.5 Questions and answers

Comments from the Tūhura’s Questions and Answers (Q&A) panel were also reviewed. Although not coded as such, enjoyment and appreciation are clear again: "This trip was the best thing EVER!! [uppercase in original] :)". “omg this place is Amazing!!”. “I thought it was absolutely AMAZING! [uppercase in original]”. “Our first visit since the upgrade. Awesome - much more hands on and the kids had a ball. Well done Dunedin ♥” “This place is the coolest”. “I loved creating art with the projector + my body - sooooo cool”. “This place is cool We stayed the night here” (age 10). “This is hands down one of the best exhibits I’ve been to Awesome job!!”. “This Place is Awesome, Took 1 ½ hours to get to This Point”. “The *Chicken Embryo* display is the coolest thing I have seen in ages, great job team!”. “This is a wonderful space for young ones to learn about science Grandpa”. “We had the best morning ever!!” (11, 12 and 35 years).

As with survey comments, Tūhura’s Q&A board contained some requests to bring back certain Discovery World’s exhibits: “You should change some of the new stuff to the mindgame [*Mind Ball*], rolling ball thing [*Kinetic Sculpture*], and the *Air Hockey* game.”

Tūhura’s Māori approach was appreciated by Māori visitors. "Nga mihi ki a koutou ma nga panni Māori [Thank you for the Māori translations] - Thank you for your signage in

[134] Children in Māori.

te reo Maori it makes things easier and friendlier for our boys". It also sparked curiosity in others. “When was the [Mā]ori language first understood?”

6.2.6 The natural world of unsolicited evidence

This thesis was defined according to a series of research questions (see Chapter 1) and the methods were carefully chosen, and instruments created to answer them (see Chapter 3). But, sometimes unexpected interactions bring about alternative possibilities (Koro-Ljungberg, 2015).

Occasionally, visitors (respondents and non-respondents[135]) approached the author to have a chat and made unsolicited comments about the science centre. The spontaneous nature of these comments is where they get their value. For example, a non-respondent elderly couple (with no children) at Discovery World approached the researcher. “Amazing… something I don’t see every day. You know, out of the daily cleaning at home and stuff” (DW, F, 83, NR), said the woman. When she was invited to see the Science Show, she got excited. “Oh, I love explosions!” (DW, F, 83, NR). Unfortunately, they couldn’t attend the show. Instead, they went to the planetarium, and when they came out, she approached again to add, “I’m getting older, but that still amazes me” (DW, F, 83, NR). She then talked about how older people need to know, both for themselves and to encourage new generations. This woman embodies the concept of science engagement and shows the potential that science centres also have to engage older generations.

Another anecdote that illustrates this potential, but also why maybe Discovery World didn’t completely fulfil it, comes from a respondent after completing the post-survey: “The only problem with Discovery World is that they should not allow children, they do not let you try anything [laughs]” (DW, M, 60s, R). It is not about banning children, but providing all generations the opportunity to engage simultaneously, as in the case of a middle-aged female respondent who was amazed by her daughter’s concentration power at the Mind Ball exhibit. The girl was happily giggling while the mom told the anecdote of how the child beat the mom.

Visitors approaching the author to share their opinion occurred more often at Tūhura. Older generations gave Tūhura a thumbs up: “What a wonderful museum. It’s terrific” (T, F, 80s, NR). “It was excellent. It was very, very good. I liked the interactives and it was nice to

[135] The quotes in this section are marked with NR when it comes from a non-respondent and R when it comes from a survey respondent (although the comment was made outside the survey). Sometimes visitors left the comments on paper, ensuring their verbatim status. When not possible, the researcher wrote their comments down as soon as possible and they are, to his best knowledge, verbatim. Comments he was unsure of are not quoted.

see so many kids with their parents. I would like to see a museum like this one in Inverca[r]gill. Currently there is none" (T, F, 69, NR). "I remember museums as static, dull, silent and boring places. Great to see what fun, learning and encouraging places they have become" (T, M, 52, R). "This new one is much, much better than the old one. Much better, so much more interactive. The new Discovery World is far superior to the previous one –very interactive and also informative. Lots of 'hands-on' things to do which keeps the children interested and engaged for over 2 hours. Thoroughly enjoyed the experience and will definitely come again. Thank you [underline in original]" (T, F, 40s, NR).

Engagement not only produces motivation to share feelings, but also information, as in the case of this young respondent after reading the scientific concepts fluency question about magnetism. "It's amazing the magnetic field is around the whole earth" (T, M, 11, R) told the author. Engagement also inspires visitors to want to know more, like a non-respondent boy, around 12, who approached the author to ask if it would be possible to add a Foucault's pendulum. He didn't know the name of the devise but described it. He saw it somewhere and, ever since, has wanted to see the rotation of the Earth himself. But this craving for knowledge does not only happen in young people. For example, a non-respondent man, around 20, approached the author after the planetarium show to ask how it was possible that the universe is expanding, if it itself contains all space.

Younger children were especially prone to expressing their impressions of Tūhura by thinking aloud. For example, a little boy, around 5, was hugging his big Elmo when the gate opened (the gate's bar was at his sight's height, blocking his vision). His eyes went big and round. "It's so beautiful!". In a different occasion, two young children, around 5 years old, were overflowing with emotion for getting to go to the butterfly house. But just when they had started running, *The Void* opened and one suddenly stopped. "Wait, this is great!". Consensus was implicit, as the whole family gravitated towards *The Void* instead of Tropical Forest. Sometimes, kids were so engaged that they had to be dragged out (literally) at the end of the visit. The screams of one of them, around 5 years old, could still be heard from the hall, until he 'escaped' and came running back to Tūhura. "He had a great time", said the mom, smiling, when she caught him just at the entrance.

The last anecdote of this section comes from Helen[136], an 80 year-old woman. "I'm 80 now and my grandma brought me to the museum when I was 6 years old. Today, I brought my granddaughters to the museum for the first time. They come from Taiwan. They came a year ago. They live now in Queenstown" (T, F, 80, NR). She wanted to bring them to the

[136] She asked that her real name be used.

museum because her grandma brought her when she was a child. “My grandma said she was tired because I was running around saying ‘Grandma!, grandma! Look at the skelentons![sic]’” (T, F, 80, NR). She explained she couldn’t pronounce ‘skeletons’ as a child, and smiled remembering how many ‘skelentons’ of animals were there. “It’s amazing how museums have changed!” (T, F, 80, NR), she added before going to sit down to rest. Later, she met another lady (in her 70s) at the bench and they discovered they come from the same group of settlers. Helen borrowed a pen and paper from the author to write down what the other lady was telling her. “She told me a lot of things I didn’t know about my own ancestors!”. Helen demonstrated several important things in a short time. First, museums can be engaging to all generations, and this engagement can be long-lasting (74 years in Helen’s instance!). Second, adults’ lifelong learning can also be influenced by visiting science centres (Gutwill, 2018). Third, museums are highly social (McKenna-Cress & Kamien, 2013; Schwan et al., 2014; N. Simon, 2010).

6.2.7 The apple doesn’t fall far from the tree

Parents’ attitudes towards science have a large positive effect on their children’s science test scores, which can come from fostering favourable attitudes towards science, including taking them to science museums (Perera, 2014).

Tina[137] is a mother of two young children, Fergus (M, 7) and Anna (F, 9). Due to his young age, Fergus couldn’t participate in this research, but Tina insisted that he do the survey and she encouraged him to fill it out with her help. Despite his survey response not being included for analysis due to the University’s approved ethics protocol for participants to be eight or older (Approval 17/062; see Chapter 3), Fergus was allowed to complete it. Parents like Tina influence their children by telling them they are smart and their parent’s help is simply a willingly accepted guide.

After the visit, Tina commented that Fergus had some questions about space travel and where he could study space science when he grew up. The author was just going to interview Dr. Oleg Abramov, a researcher from the Planetary Science Institute, about his collaboration with Space-X on the first manned mission to Mars. The institute is based in Arizona, USA, but for personal reasons the doctor was, at the time, living and working from Dunedin. The author offered to ask Dr. Abramov Fergus and Anna’s questions [138]. This led to a long conversation on space science—not only with Fergus, but with Anna also excitedly

[137] With permission to reproduce their actual names.
[138] He did it gladly. The questions and answers can be found in Appendix G.

getting involved. Both asked impressive questions, like whether Mars is big enough for all humans to live on (Fergus), and how far or close a planet can be from the Sun in order to sustain life (Anna). Tina gladly approved of Fergus' interest in becoming a scientist. By the end of the chat, Anna joined Fergus with his interest. "I want to be a scientist, too".

In agreement with other research, Tina shows anecdotally that parents' attitudes towards science can heavily influence children's own attitudes (Alexander et al., 2007; Gunderson & Levine, 2011; Halim et al., 2018; OECD, 2013b; Perera, 2014). Parents also influence in their children's learning. "Evidence indicates that the more support parents and other offer, the greater the possibility that children will learn" (Fenichel & Schweingruber, 2010, p. 77). Their effect can be capital. When learners find the real-world application of their skills and this learning is recognized and valued, they are inspired to continue learning as a lifelong endeavour (Ainsworth & Eaton, 2010).

6.3 A Place for All

The most common reason for adults to go to a museum is to bring children or grandchildren (Rennie & Williams, 2002). Discovery World seemed to be no exception, as staff repeatedly mentioned the place was seen as being only for children, with only the Tropical Forest attractive to teenagers and adults (see Chapter 2). One of the challenges in redeveloping the science centre was to make Tūhura, as a whole, enticing to all generations.

Staff think the goal was achieved. "I think we kind of flipped it from the butterflies being the heroes to Tūhura being, like, at least, equally as engaging" (DM2). "There now seems to be more teens and young adults that definitely spend more time in Tūhura, way more time than they would have in Discovery World" (SC6).

6.3.1 Science centre suitability

One way to measure if the perception about the science centre changed after the redevelopment was to ask visitors to select whom the science centre was suitable for. The three options were Children, Teenagers, and Adults, and they were allowed to select more than one.

The percentage of a group members selecting their own group[139] as suitable is here called *self-selection*. A second index, Overall opinion, was developed by the author. It is defined as the weighted opinion of all respondents with respect to a certain group:

$$\text{Overall opinion (j)} = \frac{\sum_i^4 P_{i,j} \cdot n_i}{\sum_1^4 n_i}$$

i is the respondent age group as used throughout the thesis (from 1=Children to 4=Adults), j is the age group in the specific instrument (from 1=Children to 3=Adults), n_i is the is the sub-sample size of group i, and $P_{i,j}$ is the percentage group *i* assigned to group *j*.

Visitors tended to consider both science centres as suitable for all age groups, including their own (Table 6-2). It was expected that Discovery World was going to have lower self-selection and Overall opinion in Teenagers and Adults. This counterintuitive result may be related to Tropical Forest. If these demographics found that area enjoyable, they might have been considering it as part of the whole experience.

Table 6-2
For whom is the science centre appropriate, according to visitors (N_{DW}=224, N_T=437)

	Discovery World		**Tūhura**	
Age group	Self-selection	Overall opinion	Self-selection	Overall opinion
Children	94% (61)	96%	96% (108)	97%
Teenagers	92% (33)	80%	89% (65)	84%
Adults	89% (109)	79%	84% (237)	86%

Self-selection indicates the percentage of visitors in the age group that selected their own age group. Overall opinion is how suitable the science centre was for that particular age group, but considering the opinion of all age groups. See text.

6.3.2 The significant journey from intention to interaction

The results from Table 6-2 were inconclusive, maybe because of the Tropical Forest. But there is a different way to detect differences between the centres. Before the visit, participants self-reported their intention to interact with the exhibits, see the Science Show, and visit the Tropical Forest or the planetarium. This self-report is here-on referred to as Intention. After the visit, respondents self-reported which of those activities they actually did (Interaction).

[139] For comparison purposes, this research equates Teenagers with Adolescents (13-18) and Adults with Young and Mature Adults (19+).

Not surprisingly, the difference from Intention to Interaction is not significant for the planetarium (Table 6-3). It is possible that the attendance to the planetarium could increase if it were promoted to Tūhura visitors when they were buying tickets.

> It is a shame that plan[eta]rium entrance is not part of the cost of entering discovery world [Tūhura]. Staff could explain that and ask for an extra fee or add family pass charge for plan[eta]rium. Also, I was annoyed that we were not admitted and did not want to leave and go downstairs for extra cost and ticket purchase. The young man at the door was polite and instructive. No problem there. (T, F, 53)

Anecdotally it was seen that Tūhura visitors who were unaware of the planetarium and discovered it until they were inside Tūhura, seldom went out to buy tickets and return. This is probably because the 'cost' of going back and forth is higher than the perceived benefit (Bitgood, 2016). Once visitors walk towards the exit, even for reasons other than leaving, the probability of their leaving increases (Falk & Dierking, 2016).

The most conspicuous result is that visitors tend to do more than they expected to, especially with the exhibits, where Interaction doubled from Intention at both science centres (Table 6-3).

Table 6-3
Intention of doing activities and actual Interaction

		Discovery World		Tūhura	
Tropical Forest	Intention	90% (202)	p=.002, OR=5.00,	76% (764)	p<.001, OR=11.0,
	Interaction	97% (218)	d=0.89	94% (944)	d=1.32
	Difference in Interaction "N-1" $\chi^2(1)=3.10$, p=.078				
Exhibits	Intention	41% (91)	p<.001, OR=23.0	46% (463)	p<.001, OR=37.5,
	Interaction	80% (179)	(d=1.73)	93% (937)	d=2.00
	Difference in Interaction "N-1" $\chi^2(1)=33.0$, p<.001				
Science Show[140]	Intention	17% (39)	p=.014, OR=1.91,	NA [a]	
	Interaction	27% (60)	d=0.36		
Planetarium	Intention	NA [b]		19.5% (196)	p=.351, OR=1.56,
	Interaction			20.9% (210)	d=0.242

[a] There were no Science Shows at Tūhura.
[b] The planetarium had a separate entrance at Discovery World.

Interaction with exhibits at Discovery World increased steadily (first time visitors (78), 78%; two-three visits (43), 81%; four plus visits (58), 91%) and significantly (Pearson's

[140] Note that the relatively low numbers of Science Show attendees is greatly influenced by the availability of the show. While the Tropical Forest and the exhibits were open at any moment from 10am to 5pm, the Science Show was only available on weekends and holidays twice a day (11:00am and 3:00pm). The show lasted for half an hour, which meant that it was available for 1 out of 7 hours. That a quarter of the population saw the Science Show, when its hours were so restrictive, is quite impressive.

$\chi^2(2,224)= 7.904$, $p=0.019$, $V=0.188$). This indicates engagement builds up over visits. However, it required at least four visits to surpass the 90% threshold. In contrast, Tūhura's exhibits reached 94% (573) Interaction from the first visit and remained around this value in subsequent visits.

Lastly, notice the drop in Intention to visit Tropical Forest, from 90% to 76%. It is not a negative finding—rather, the opposite. Visits to Discovery World were almost exclusively driven by the Tropical Forest, but Tūhura does not depend solely on its attraction power anymore; more people visit Tūhura mainly because of the exhibits. Discovery World's exhibits were mainly popular among Children. Interaction by visitors over 19+ was as low as 76% (93). Tūhura unified Interaction across all ages, with all age groups interacting over 90%. Science exhibits at Tūhura were not less popular than the Tropical Forest. Both exhibits and Tropical Forest enjoy a similar level of visitor Interaction, almost reaching saturation.

6.3.3 "This is a great place for all ages"

Evidence of Tūhura being suitable for all ages was found in survey comments: "This is a great place for all ages" (T, F, 10). "It is great that I as an adult could [e]njoy the exhibition as well as there was a level [unintelligible] stuff if you are interested" (T, F, 26). "Really enjoy brin[g]ing my kids here they enjoy it but so do I as a[n] adult. Two birds one stone. And we all learn someth[i]ng new" (T, F, 27). "Had an amazing time. Not sure who enjoyed it more the adults or the children" (T, F, 34). "I always come here when we are down from [Christchurch]. I love this place and want to live here one day" (T, F, 42). "I think that this is a very interesting, informative, interactive area for young and old. I thoroughly enjoyed bringing my grandchildren here today. Many thanks" (T, F, 56).

6.3.4 "I wanted to come since it opened"

As conclusion for this section regarding whether or not the redevelopment made Tūhura attractive to adults as well, the following anecdotal evidence is provided.A non-participant woman in her 70s, as she was leaving, made an unprompted comment to the author. "Thank you. That was lovely. I wanted to try them all!"[141]. She was accompanied by her three daughters in their 40s and 50s. No children were part of the group.

The mother and one of the daughters stayed on a bench outside Tūhura while the other two ran an errand. The researcher approached them to introduce himself and ask for the

[141] These comments were written down by the researcher in situ. The best effort was made to transcribe them verbatim.

permission to use the quote. This led to a long chat. It was challenging to capture verbatim sentences because they were excitedly laughing and speaking at the same time. That's probably the main takeaway—an indescribable sense of joy and engagement was visually transmitted with their laughter and facial expressions.

The daughter described Tūhura as "fun, interactive, and interesting", and the mother described her reaction to the Tropical Forest as "ecstatic", "I would tell everyone to come, you are missing out. It's worth going". They loved that the new slide is now for adults as well, not only children. They both used it twice. "Twice and I didn't get stuck!", said the mother, laughing. "Maybe you should keep coming until you get stuck!", replied the daughter in the middle of a laugh. They even offered to take care of the iPads so that the author could go try it for himself and see how fun it is. The daughter said they had wanted to take their mother for a while, because she was so keen to come. The mother confirmed it. "I wanted to come since it opened".

6.4 Back to the Gap: putting gender aside and bringing engagement in

Previous chapters showed a gender gap in self-concept in science (see Chapter 5) and scientific knowledge (see Chapter 4), both favouring males. This chapter in science engagement is mainly qualitative and therefore not suitable for inferential statistics. There were only two quantitative factors in engagement. One was visit time, but the fact that visitors come in groups avoids the possibility of analysing it by gender. The other one was the engagement scale, which didn't show any gender differences, probably because of a ceiling effect (see Section 1.2.2). Still, it is possible that engagement is more closely related to gender gaps that it may appear.

6.4.1 The evident, yet elusive, gender gap in science

The literature around a gender gap in science seems at first to be a bit contradictory, but this may not be so. According to Riegle-Crumb, Farkas, and Muller (2006), gender differences in advanced-level math and science course-taking have been virtually disappearing since the mid-to-late 1990s. Even for fourth and eighth graders, gender does not create a large number of significant differences in students, internationally (Mullis et al., 2008). "[W]e do not find evidence of a consistent white male advantage in entrance into STEM postsecondary fields" (Riegle-Crumb & King, 2010, p. 660). For the specific case of New Zealand, PISA does not report a gender gap in science (OECD, 2009, 2014).

Gender disparities in high school academic indicators do not explain differences in the choice of a STEM major (R. M. Simon & Farkas, 2008). But despite the evidence about females not being at an academic disadvantage, underrepresentation of girls and women in some areas of science is still a well-known phenomenon (A. J. Friedman, 2008; W. M. Williams & Ceci, 2012).

In Chapters 4 and 5, a gender gap in scientific knowledge and self-concept in science was reported. To have a better idea of the gap and why some research doesn't find one, prior self-concept in science[142] and scientific knowledge in Tūhura were recoded from 0% (lowest possible value in each case) to 100% (largest possible value). Then, for each age cohort, the female median was subtracted from the male median[143]. This difference is the gender gap in percentage (Figure 6-2)[144].

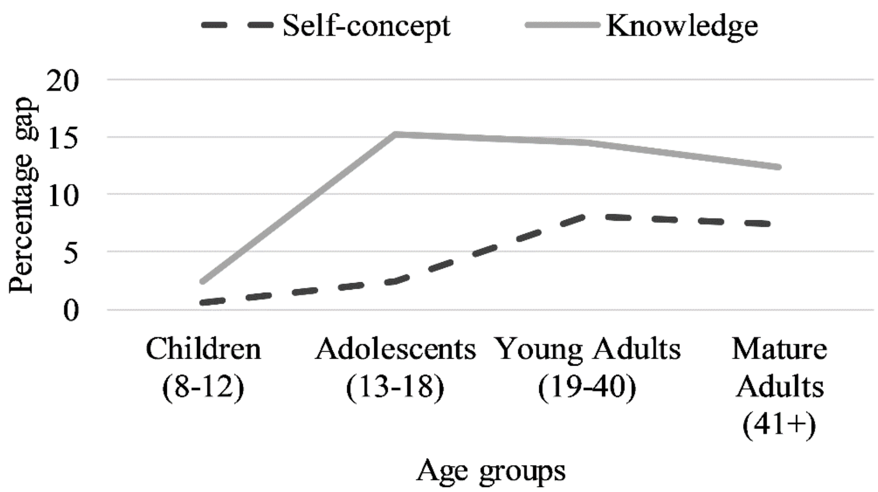


Figure 6-2. Percentage gap in scientific knowledge and self-concept in science (males median minus females median).

The first thing to note is that there is no gender difference in Children for either self-concept or knowledge. The gap in knowledge dramatically increases from Children to Adolescents, remaining relatively stable in adulthood. However, the gap in self-concept does not appear until adulthood.

142 Data from the Likert scale were used, as its sample size was larger than that from the Visual Discrete Scale.
143 Since some groups are small (up to only 20 in male adolescents), median instead of mean is used.
144 If it is positive, it means it is in favour of males. Negative would mean in favour of females. Around zero means no gap.

The reason why an assessment like PISA does not detect a gender gap in science may be because it only assesses 15 year old students[145]. This is the age when boys start pulling ahead in IQ score[146] (Flynn, 2012; Lynn & Irwing, 2004) and the gap in STEM starts becoming evident (Mostafa, 2019).

Why the gap would apparently emerge first in scientific knowledge, rather than in Self-concept may be related to several things. First, Figure 6-2 does not represent the temporal evolution of Knowledge and Self-concept in a group of individuals as they grow up, it shows snapshots of these two constructs in individuals coming from different generational cohorts. As explained in section 3.3.5, people of different cohorts grew up under different paradigms and therefore the evolution of their gaps may as well be different. In other words, it is possible that, for example, a gap in self-concept appeared prior to a gap in knowledge in Mature Adults (41+) when they were younger, and that Children won't ever develop a gap in either construct. It is not possible to make absolute claims about the evolution of the gaps in each cohort without decades-long longitudinal research. Notwithstanding, it is possible to turn to the literature to propose a possible shift in the paradigm of gaps appearance that goes accordingly to cohort differences. The author proposes that the gap began in self-concept in older generations, but begins in knowledge in younger generations. In Adolescents, the gap would actually not be in knowledge in first instance either, but in engagement. The less engagement with science in girls would lead to less study in science related topics, which would lead to a gap in scientific knowledge. The gap in self-concept wouldn't be unrealistic (as it might have been in Adults); all the opposite, it would be a fair self-assessment of lesser knowledge. The following subsections are aimed at justifying and clarifying this proposal.

6.4.2 A matter of confidence?

Achievement and self-confidence are constructs that mutually reinforce one another (DeBacker & Nelson, 2000; Huang, 2011; Jansen et al., 2015; Jansen et al., 2014; Marsh & Martin, 2011; Marsh et al., 2012; Wender, 2004; Wilkins, 2004). However, whether high performance promotes self-confidence or if it is high self-confidence what promotes performance is an unresolved and unsolvable dilemma (Meier & Diefenbach, 2018).

[145] Given that gaps can change rapidly in a few years, this research also acknowledges grouping respondents into age cohorts may conceal details of the trends.
[146] There is evidence that a fully matured man has an IQ 5 points higher than a fully matured woman. Nevertheless, the reasons are not genetic. For a full discussion, see Flynn (2012).

Although the relation between self-concept and achievement may be bidirectional, a common explanation for the gender gap posits that the gap in science originates in self-beliefs, e.g., females underestimating their ability in physics compared to male students (Jansen et al., 2015). This would possibly be due to stereotypes. “[A]lthough students, especially girls, refute bald statements of stereotypic beliefs, they may still be influenced by cultural stereotypes on a less conscious level” (DeBacker & Nelson, 2000, p. 251). For example, the common stereotype of a scientists is white male (Fenichel & Schweingruber, 2010).

In such a scenario, constructs like self-concept and self-efficacy would play an important role in explaining gender differences in interest in science (Krapp & Prenzel, 2011). If so, closing the gap in STEM would probably entail closing gaps between gendered self-concepts and perceptions (Lee, 1998). However, Figure 6-2 doesn’t support this explanation; here, the gap emerges first in scientific knowledge, not in self-concept in science.

6.4.3 The role of engagement in learning

Tūhura visitors were asked to rate from one to five stars the Tropical Forest, the exhibits, and the planetarium in terms of fun (an important component of engagement) and learning (Figure 6-3). All attractions, and especially the exhibits, were highly rated for both fun and learning. Moreover, for the exhibits, fun and learning were highly correlated (Wilcoxon-Pratt test, $Z=8.73$, $p<.001$, $r=.469$, 170 discordant pairs of 347).

Structural Equation Modelling (in SPSS AMOS v25) was used to find correlations among prior self-concept (Visual Discrete Scale format), engagement, and fluency[147] (Figure 6-4). Ellipses represent the constructs, rectangles the individual items, and circles the inherent item errors. The three constructs turn out to be inter-correlated. From separate data, self-concept in science (Likert scale format) was also found to be significantly correlated with scientific knowledge (Pearson’s correlation, $r=.360$, $p<.001$, $N=438$).

[147] Recall that fluency is a measure of self-reported knowledge (see Chapter 5).

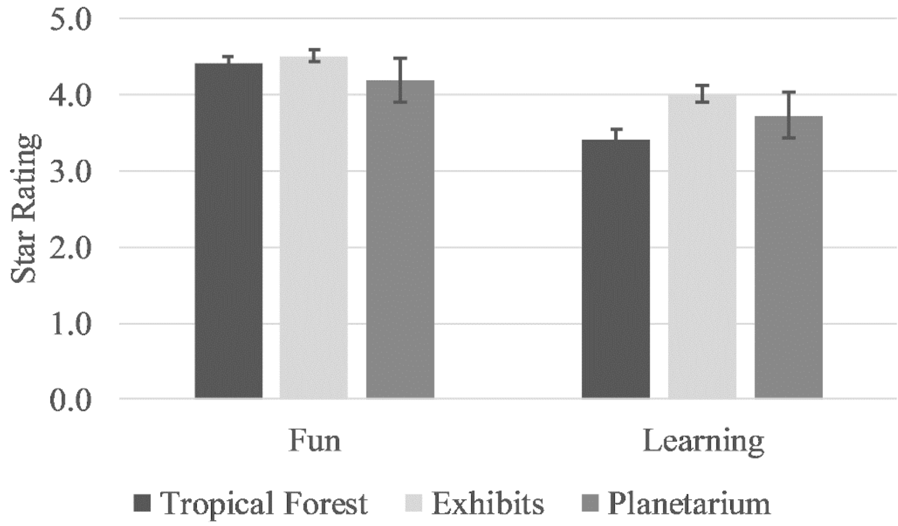


Figure 6-3. Star-rating means for fun and learning at Tropical Forest (n=336), exhibits (n=347) and planetarium (n=65). Confidence Intervals at 95%.

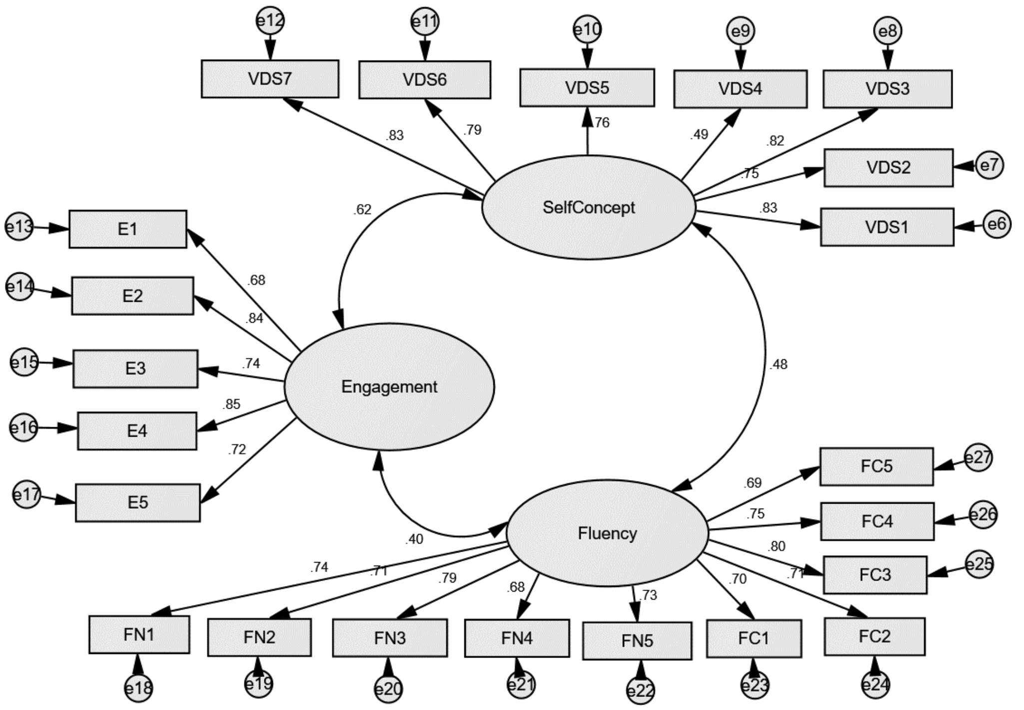


Figure 6-4. Structural equation modelling with self-concept in science (seven items, Visual Discrete Scale), science engagement (five items) and scientific fluency (ten items).

So far, it has been shown that science engagement is correlated to self-concept in science and scientific knowledge. What has not been shown is how it can be involved in gender gaps in those constructs. For that, let's go down a different road. In Chapter 4, it was shown that whether one read the exhibit panels (or not) did not influence learning (Table 4-

3), but those who read panels had prior scientific knowledge (M_R=2.08, n_R=312)[148] significantly higher (t(425)=2.63, p=.009, d=0.127, d_{CI}=0.109) than those who did not read the panels (M_{NR}=1.71, n_{NR}=115)-. This contradicts Diamond (1986), who suggested that most visitors only read labels if they have trouble figuring out what to do, as that would imply panel readers were less scientifically literate than non-readers.

Prior self-concept in science in panel readers (M_R=3.76, n_R=306) is also higher (t(418)=2.15, p=.032, d=0.105, d_{CI}=0.110) than in non-readers (M_{NR}=3.57, n_{NR}=114). Also visit time (which can be used as a measure of engagement) may be[149] longer (t(427)=1.95, p=.051, d=0.094, d_{CI}=0.108) in panel readers (M_R=70:59min, n_R=312) than in non-readers (M_{NR}=64:58min, n_{NR}=117).

Before accepting that these data contradict Diamond, the effect of age was checked. This is because adolescents and adults have higher level of prior scientific knowledge than children (see Chapter 3). Since children are less likely to read labels than adults (Diamond, 1986; Serrell, 2015), age was considered a possibly confounding variable[150]. However, this possibility was ruled out because of two facts. First, Children represent the same percentage of the demographic (26%) in both readers and non-readers. Second, age can't explain the correlation between reading and not reading panels with prior self-concept in science[151] (see Chapter 5) and with visit time[152] (see Chapter 3).

An alternative to Diamond's claim comes from considering that panel readers can be divided into those who read most of a label and those who read very little (Falk & Dierking, 2016). Those who read very little would be those who only read to find instructions, while those who read the most are more engaged. Engaged visitors don't necessarily read to find instructions, but to know more about the phenomena they are witnessing. If this is the case, it can be expected that the latter type of panel-readers like to be informed about science beyond what is offered at the science centre. Previous readings, documentaries, talks, etc. would have increased their scientific knowledge.

Their higher level of engagement would explain their longer visits. Also, if we consider their self-concept assessment as fairly accurate, their self-concept in science would increase as their scientific knowledge does.

[148] NR stands for visitors that did not read panels and R for those that read panels.
[149] This is an example where p-value can be tricky. It is so close to α that it can be considered 'significant', but the effect size is small and the confidence interval of the effect size is large. A larger sample size would be needed to know if that p-value is a type-I error.
[150] An extra variable having a hidden effect.
[151] Self-concept is not correlated with age.
[152] Visitors come in groups with mixed ages. All members of a group have the same visit time.

6.4.4 The hypothesis of the river delta

STEM careers can be divided into two broad categories, physical STEM careers and life sciences STEM careers (Mohtar et al., 2019). It is well documented that girls tend to have less interest in 'hard' sciences than boys (M. G. Jones, Howe, & Rua, 2000; Krapp & Prenzel, 2011; Labudde, Herzog, Neuenschwander, Violi, & Gerber, 2000; J. Osborne & Dillon, 2008). Females tend to be more attracted to biology and males to physics (Akarsu & Kariper, 2013). In these choices, engagement is a cornerstone, as it supports effective science learning and interest in learning more (Ainley & Ainley, 2011; Csikszentmihalyi et al., 1997; Krapp & Prenzel, 2011; McCallie et al., 2009; Renninger & Hidi, 2002).

However, career choices are influenced not only by confidence and interest in science, but by relative academic strengths (Stoet & Geary, 2018), and it was found in 2015 PISA that boys had a significantly larger rescaled intra-strength[153] in Science, while girls' intra-strength was in Reading (Stoet & Geary, 2018). Girls may tend to take fewer science and math courses during secondary school (Berryman, 1983; Maple & Stage, 1991) and be underrepresented in STEM not only due to lower self-beliefs in understanding science, but due to having more areas where they feel they can succeed (Mostafa, 2019). To close the gap, tackling boy's underperformance in reading may be just as important as supporting girls in pursuing science-related careers (Mostafa, 2019).

An approach that is consistent with the results of this research is that the primary factor for women's underrepresentation is not discrimination or ability, but choices that start at a young age, both made freely as well as constrained by biology and society[154] (W. M. Williams & Ceci, 2012).

Based on all the above, the hypothesis of the river delta is that girls choose non-physical STEM careers not because they feel they can't or because they are not allowed to study them, but because "[u]nlike many high-performing boys, many high-performing girls may not pursue a career in science, even if they are capable of succeeding in it, because they are likely to be top of the class in non-science subjects too" (Mostafa, 2019, p. 5). Thus, when their capacity allows them to *choose* from several options, they are inclined to go for biology-based subjects. Although not completely free, it is still a choice, and it is based on both options and level of engagement with those options.

[153] PISA assess three main subjects: Science, Reading and Mathematics. While there wouldn't be a gap in Science in absolute terms, boys tend to score higher in Science than in the other two subjects, and girls do so in Reading. Since each area had different number of items, they were first rescaled.

[154] For example, maternity can be a free choice, but it is ruled by biological and societal restrictions (W. M. Williams & Ceci, 2012).

More than an 'science engagement gap', it can be viewed as an engagement in a different field[155]. Choosing a stream other than science is what would produce a gap in scientific knowledge, and the ideal self-assessment would reflect this gap in the form of a gap in self-concept in science[156].

6.4.5 Generational shifts

If gender differences are explained by engagement differences, and engagement drives learning, and if reading science exhibit panels is an indirect measure of science engagement, then it should be possible to see some kind of resemblance of these gender differences between panel readers and non-readers. Table 6-4 shows results from Mann-Whitney U tests conducted before the visit on scientific knowledge and self-concept in science depending on gender (males / females) and whether or not visitors read some panels.

Table 6-4
Comparisons of knowledge and self-concept in different age groups by gender and panel reading

	Gender		**Panel reading**	
Age	Knowledge	Self-concept	Knowledge	Self-concept
8-12	.618	.819	.466	.494
13-18	.005	NA	.028	.046
19-40	.001	.003	.012	.089
41+	.008	.006	.395	.819

Mann-Whitney U test was conducted before the visit for Gender (males / females) and Panel reading (readers / non-readers). The reported figures are p-values.

The first three age cohorts behave as expected. The results for Adults (41+) may initially seem counter-intuitive, but are explainable. The author agrees that lower self-beliefs in science amongst females was the origin of the gap when it was first studied; however, societies evolve—what was true last century may not apply to new generations.

The theorized explanation is that the gender gap in older people (41+) indeed originated in lower female self-beliefs, but the source has been changing. Herd (2005) reviews the situation of women in New Zealand, finding efforts to make it a more equitable society in the last few decades. Even though work still needs to be done, changes have

[155] Otherwise we should also talk about engagement gaps in boys in all the fields they don't choose. What boys don't choose is as important as what girls don't choose. Fields where females are more interested than males are likely to have a knowledge gap in favour of females.
[156] It is not claimed that this is the only reason, but an alternative with probably at least the same weight as the traditional low self-beliefs explanation.

become evident. “In recent years girls have performed better at school[157] than boys, causing some head scratching among educators and parents” (Herd, 2005, p. 82).

As a reflection of New Zealand (to some extent), we can look at Australia, where it is reported that the proportion of females studying science at the University of Western Australia increased from 18% in 1968 to 60% in 2008, and from 1% to 17% in engineering. This is partially due to programs like the Women in Science and Engineering Program (WISE), established in 1989 (Longnecker & Davis, 2013).

If, as it is theorized, the origin of the infamous gender gap in science is cohort-dependant, then graphic representations like Figure 6-2 need to stop being seen as predictive. In 50 years, the shape of the graph will not be the same. There are multiple initiatives to bring females to science starting when they are young. This is completely laudable, but they may not be producing all the results they could. For example, the U.S. female workforce in STEM tripled from 7% to 23% in only two decades, from 1970 to 1990. But over the next two decades, from 1990 to 2011, it only grew to 26% (United States Census Bureau, 2013). Might this have something to do with current initiatives targeting gap sources from several decades ago and not targeting the main current sources?

6.4.6 Closing the engagement ‘gap’

As stated above, it is about more than closing a gender gap. It is about making science as interesting to females as other fields. The science centre showed an interesting change in this respect.

In terms of Intention and Interaction (see Chapter 2), males came to Discovery World to play with the exhibits (51%, 43) significantly more (‘N-1’ $\chi^2(1)=5.64$, $p=.018$) than females (35%, 48). This difference is consistent with the self-reported Interaction (‘N-1’ $\chi^2(1)=7.66$, $p=.006$; males=89%, 76; females=74%, 103).

Tūhura is different. Intention in males (50%, 198) and females (44%, 261) was not statistically different (N-1 $\chi^2(1)=2.90$, $p=.088$), but still very close to α. However, Interaction was not different at all (‘N-1’ $\chi^2(1)=1.10$, $p=.294$; males=94%, 377; females=93%, 549). Children were the only cohort where boys still had a statistically larger tendency (N-1 $\chi^2(1)=8.69$, $p=.003$) for Intention to play with the exhibits (66%, 89) than girls (49%, 64). However, this difference disappeared for Interaction (‘N-1’ $\chi^2(1)=2.22$, $p=.137$, males=96%, 128; females=91%, 120). These results show that Tūhura, unlike Discovery World, is presenting science as an interesting subject, regardless of gender.

[157] This is a general statement, not specifically for science.

6.5 Conclusions

This chapter focused on engagement as the third element of scientific literacy. Evidence suggests that visitors of both science centres arrived with a prior high level of science engagement. However, science centres have the capacity to increase this engagement even more. For example, visitors do more than they expected. Self-reported Interaction after the visit was double that of the pre-visit Intention to play with the exhibits at both centres.

Although both centres were effective at engaging visitors with science, Discovery World depended more on the Tropical Forest. Exhibits required four or more visits for visitors to reach a level of Interaction with the science exhibits close to their Interaction with the Tropical Forest (over 90%). In contrast, Tūhura's science exhibits reached a comparable level to the Tropical Forest in the first visit. The quick engagement power of Tūhura's exhibits is especially important for visitors who do not have the opportunity to come back.

One of the main goals of the redevelopment was to make the science centre engaging to all generations. Results clearly indicate that this goal was met. Not only did adults bring their children to enjoy the experience, adults with no children were now attracted to the exhibits.

A gender gap in scientific knowledge (Chapter 3) and self-concept in science (Chapter 4) was also analysed. Evidence suggests that a common origin for both gaps could be a personal choice, rather than external restrictions and low confidence. Girls' choice would be, in-turn, related to having several career options and being especially engaged with non-science subjects, rather than with science. If so, closing the gap would be more related to engaging girls with science than with convincing them that they can be whatever they want to be. Interestingly, Discovery World was more engaging for males than for females, but Tūhura is equally engaging to both genders, which may mean that it is indirectly helping to close the gender gap in science.

The focus of the next chapter is crucial for the development of modern science centres: understanding what exhibit characteristics encourage everybody to get engaged with science regardless of age or gender.This chapter finishes with a maxim expressed by a Tūhura visitor that can work as the preamble of Chapter 7: "Science like life is an experience that is enjoyed most when being active and having fun" (T, M, 35).

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

7.1 Introduction
7.2 Exhibit Examples
7.3 Exhibit Characteristics Related to Attraction Power
7.4 Exhibit Characteristics Related to Holding Power
7.5 Aspects Related to Learning Power
7.6 Conclusions

Chapter 8:
CONCLUDING REMARKS

Chapter 7: EXHIBITS AND SCIENCE LEARNING

7.1 Introduction

In previous chapters, learning was defined as a change in scientific literacy. Research at both Discovery World and Tūhura helped show that a single visit to a science centre can significantly increase scientific literacy. But this doesn't mean all science centres and all exhibits have the same capacity to foster science learning. "Everything around the object has an impact on how the visitor reacts, interprets and assimilates information" (Gazi, 2014, p. 5). This chapter describes and analyses exhibit characteristics that promote science learning. Shettel et al. (1968) proposed a series of seven variables related to exhibit effectiveness: Ability to attract attention, Holding power, Ability to bring about a change in level of interest, Ability to bring about a change in attitude, Change in level of concept knowledge (recall), Change in level of factual knowledge (recall), Change in level of multiple choice knowledge (recognition), and Change in level of exhibit-specific knowledge. These variables were reduced to three main areas for consideration here: Initial attracting power, Holding power, and Teaching effectiveness. Bitgood (2016) has called them Attracting power (initially attracting visitors to the exhibit), Holding power (maintaining the attraction) and Learning power (optimizing relevant learning).

An outcome of paying attention in a museum setting is learning. While the process of paying attention is a continuum, it can be analysed in three main stages: the Capture Stage (initial attention capture), the Focus Stage (attention narrows down to a specific item for a few seconds) and the Engage Stage (visitors become deeply involved with the exhibit content) (Bitgood, 2010, 2016).

Attraction power has the largest effect on the capture and focus stages. Holding power is more directly related to the Engage Stage. Learning power is the product of sustained attention, and it is a cumulative effect of all stages. It is not enough that an exhibit has a culminating point; "*every intermediate step* [emphasis in original] in the visitors' experience must be sufficiently motivating that they make the choice of continuing to invest time and attention there" (S. Allen, 2004, p. 18).

SWOT focus groups, interviews and visitor comments were the sources analysed in search for exhibit factors that promote science learning through attracting, holding and learning powers. The found characteristics are mainly related to scientific knowledge and science engagement. Data from staff interviews and visitors comments were coded according to what was described in section 3.5.2. Exhibit characteristics were placed in the Power section where they have the biggest influence. Supporting literature is added to corresponding characteristics. When applicable, these findings lead to recommendations one should bear in mind during the design or selection of a new exhibit. Focus groups. To facilitate quoting, each source was assigned an abbreviation (Table 7-2). The characteristics that will be discussed throughout this chapter are summarized in Table 7-3.

Table 7-1
Sources of data used in analysis of exhibit characteristics

Source	**Method**
Museum staff	Interviews
Highly engaged children and adolescents	Focus groups
Science centre visitors (respondents)	Surveys
Science centre visitors (non-respondents)	Informal chats

Table 7-2
Abbreviations used to specify the source of quotes

Abbreviation	**Meaning**	**Examples**
(SCX)	*Science Communicator* number *X*.	"Yes, science communication can be hard, but it can be fun too" (SC4) is a quote from the fourth science communicator.
(DMX)	*Decision Maker* number *X*.	(DM1) means first decision maker.
(DW/T, F/M, Age)	*Discovery World* / *Tūhura* respondent visitor, whose gender is *Female* / *Male* and is *Age* years old.	(DW, F, 8-12) means female Discovery World visitor in the range of 8 to 12 years old. (T, M, 42) means 42 year-old male Tūhura visitor.
(DW/T, F/M, NR)	*Discovery World* / *Tūhura*, non-respondent visitor (didn't participate in the survey), whose gender is *Female* / *Male*.	(T, M, NR) means Tūhura visitor, male, non-respondent.
(DW/T, C/A)	Participant(s) from the *Children* / *Adolescents focus group* carried out at *Discovery World* / *Tūhura*.	(DW, C) means child (or children) from the Discovery World focus group.

Table 7-3
Exhibit characteristics related to attraction, holding and learning powers

Attraction power	Colour and other stimuli
	Visibility
	Transferable attraction
	Novelty
	Intergenerational interaction
	Motivation to read panels
	Spatial layout of exhibits
Holding power	Enjoyment
	Comfort
	Interactivity: property of the exhibit
	Interactivity: the user as part of the exhibit
	Social interaction
	Challenges
	Testing
	Diversity of topics and exhibits
	Linked concepts
	Lighting
	Design and maintenance
Learning power	Understanding what science is
	Inspiration and passion
	The science in science centre exhibits
	Phenomena exposure
	Immediate apprehendability
	Instructive labels
	Learning, not just engagement
	The telephone game
	Personal relevance
	Post-visit engagement

7.2 Exhibit examples

Several exhibits from Discovery World and Tūhura are referred to throughout the following sections as they exemplify characteristics of aspects related to learning science. These exhibits were hand-picked by the researcher and are described in Appendix C. Selection was based on three criteria: mentions by focus group participants, mentions as favourite exhibits by museum staff and visitors (see Appendix C), and mentions by survey respondents when they answered what they learnt (see Chapter 4).

When there were multiple exhibits that could be chosen as examples, preference was given to exhibits related to light and electromagnetism because of this thesis' focus. Some exhibits were explicitly mentioned as not good for learning, or they were completely off all participants' radars. "'The least favourite'. The ones that you don't really remember" (DM1). Those were picked to specifically show areas for improvement. As during previous chapters, exhibit names are italicised for sake of identification.

7.3 Exhibit Characteristics Related to Attraction Power

Catching a visitor's attention is the first stage in the process of learning at a museum (Bitgood, 2010, 2016). Attracting power is therefore crucial to setting the scene for learning. It refers to what makes a visitor approach a certain exhibit and stop for at least a short period of time (Serrell, 1998).

7.3.1 Colour and other stimuli

Usually, humans first interact with their surroundings visually. An exhibit's appearance or observable features are amongst the first things that grab visitors' attention (DeWitt & Osborne, 2010).

Extreme or unusual colours are deemed 'fun' because they depart from the traditional school science approach (Appelbaum & Clark, 2001). Focus groups stated that exhibits should include a good selection of several colours, especially ones that glow. The best example of glowing and varied colours is the *Plasma Room*: "I think the *Plasma Room*'s like a really good example of where the Light Room [the Light Zone] should be heading type of thing" (DW, A).

Tūhura's *Coloured Shadows* is a colourful example, but still not attractive. It needs "something that makes you look at it and wonder 'What is that?'" (DW, A). Noise, larger size, contrast, and movement or bright lights can be used to attract attention. Our survival reflexes make us to pay more attention to these properties (Bitgood, 2016).

Focus groups suggested how a static exhibit can become attractive by adding movement. When nobody is using *Coloured Shadows*, the lights could be moving around automatically, like chasing targets. Moving objects are more attention-getting than static objects (Bitgood, Patterson, & Benefield, 1988). Once a visitor's attention has been grabbed and they start interacting, lights can return to their usual position.

Similar principles apply to panels. For example, outside *The Void* there is a sign on the floor. Adolescents noticed it because "it's a nice bright colour on a dark floor, so it stands out more, so easier to spot that" (T, A).

7.3.2 Visibility

No particular desirable feature will have its expected effect if visitors don't see it. Visibility can be diminished in different ways. The easiest to understand is physical barriers, like walls. Unseen Forces Zone has this problem.

> I reckon there are too many walls in all of the light exhibit [Unseen Forces Zone], because most of them you couldn't see. They were hard to see and find, so you kind of had to have to walk around a lot … Quite a few of the exhibits were hidden. (T, C)

Being hidden also causes poor visibility. The *Plasma Room* was technically part of the Light Zone in Discovery World. But "people tend to go into the *Plasma Room* more because it's right there, like in your face, you see it" (DW, A). While the Light Zone was "just the hole in the wall that you walk into" (DW, A). Lack of salient elements in the exhibits can make visitors take paths where they don't pass by certain exhibits and end up missing them (Bitgood, 2016).

> *The Void* is a massive part of the whole Tūhura, but I did completely pass it without realising the first time I went to Tūhura since it was revamped. I think the whole area is pretty. It's very interesting and it's a shame that it's kinda quite discrete and a lot of people ─not ignore it─ but go right past it, without doing it justice. (T, A)

Another cause of invisibility is distraction. If something else has a high attracting power, attention is diverted and the relative visibility of other close exhibits is reduced. Visitors are drawn to the most visually-compelling exhibits and those with intrinsic interest to them (Falk & Dierking, 2016). Focus groups noticed this at Tūhura. They mentioned a visual path that runs from the entrance to the *DNA Slide* and adjacent exhibits (the most illuminated and colourful part of Tūhura), bypassing the Unseen Forces Zone (the darker space). "People are gonna walk towards the things like the *Slide* and the skeleton [*Skeleton Bike*] that are just really obvious and look interesting. And then the lights [Unseen Forces Zone exhibits] are off to the side behind walls" (T, A).

Figure 7-1 is a photograph taken from the entrance of Tūhura. The *DNA Slide* and *Skeleton Bike* were more isolated and brightly illuminated. Isolation of exhibits by lightning and position may lead to the perception of higher significance and value (Schwan et al., 2014). Although the investigation by Eghbal-Azar, Merkt, Bahnmueller, and Schwan (2016) is not on science exhibits, but a museum displaying German literature, they found that position and size of exhibits play an important role in selection processes. In their research at zoos, Bitgood et al. (1988) found that large size and unusual appearance are features that can be perceived at a distance and attract visitors. The *DNA Slide* has all these features—it's isolated, brightly illuminated, big, and unusual in appearance. It's not surprising that it pulls visitors to it.

Figure 7-1. Photograph taken from the entrance of Tūhura. The Unseen Forces Zone (left) is darker than the zone with the *DNA Slide* and the *Skeleton Bike*.

The visual path leading to a particular zone in Tūhura may be having an effect on what visitors see and don't see. For example, although the planetarium is still in the same place it was before the redevelopment, several visitors reported missing it.

> I missed the planetarium and only saw it on the way out. I guess I was attracted to the lights and sounds coming from the other room. Maybe the entrance needs more lighting or something to draw your attention. (T, F, 40)

Panels also must catch a visitor's eye in order to be read. For this to happen, position is paramount. The first example comes from *The Void*. There were two signs before entering, one beside the door and the other on the floor. Many focus groups participants didn't notice the one by the door. This may be because people tend to focus on objects in their line of sight (Wineman & Peponis, 2010) while those beyond it may remain completely undetected (Bitgood, 2016). But, the sign on the floor is not much better. "People aren't likely to look at the ground" (T, A). The solution would be to put it on the door.

> Some idiots, like me, they don't tend to read signs that are on the floor. So, maybe on *The Void* one, you should put the sign on the door, instead of on the floor. Because there was a sign on the floor that said 'If the light is red, don't go in'. The light was red, I didn't see the sign and I walked in. (T, C)

The previous quote highlights how a simple unread sign can affect other visitors' experience. The following conversation now shows how it can affect their own experience. It happened at *Frozen Shadows* (DW, C).

> ─People going in there [*Frozen Shadows*] and once we were in there, they asked us how it works. And there were instructions. Right there.
> ─Literally, there should be a 'Please read sign' on it. Because they just don't read it. They just go in and they don't know what they're meant to do, except press the button.
> ─They're far away from it [the wall] and then they just 'What happened? I don't know how to do this, it's not working'.

However, while the Children's focus group found *Frozen Shadows* label easy to see, members of the Adolescents' focus group (and the author himself), didn't see that panel. Strangely, the opposite happened in Tūhura. Some panels that were clear for the author and adolescents went unnoticed by the children.

Height is a possible explanation for why adolescents see/miss what children miss/see. "Some of the signs are too high or too low" (T, C). A label's height determines whether or not the visitor's line of sight coincides with the text (Serrell, 2015). The label's proximity to the exhibit is also critical. Labels close to the object capture attention; if the label is not immediately close, it requires more effort and can be missed (Bitgood, 2016). More research is still needed to understand what effect a label's height truly has, or if there was a collateral factor that produces an effect of labels being 'hidden in plain sight'.

Little can be done to have panels corresponding to every visitor's height, but there are other ways to make them visible. Unlike *The Void, Frozen Shadows* didn't have a door to place the label on, but "the sign, the instructions should be on the wall as you walk in, so that you can see what you are supposed to do as you walk in. So, it captures your attention" (DW, C). Visitors would like information adequate for children and adults alike (see also 7.5.6). "I would like two levels of info - basic and higher as I love science and always want more in depth info" (T, F, 36). One possibility could be having labels with text at eye level for adults and large graphics at a child's eye level (Serrell, 2015).

Position can also assign importance to a label. There is a big panel with explanations on the electromagnetic spectrum at Tūhura. Adolescents in their focus group found it a good addition, easy to understand, and interesting, but most of the Children didn't even notice it. This time, the difference wasn't because of height. First, the label doesn't have an associated exhibit. Second, it is on the back of another exhibit (*Monochromatic Room*), and visitors mostly ignore exhibits that are obstructed or off side (Wineman & Peponis, 2010). Due to those characteristics, Children may have unconsciously assigned low value to this panel and skipped it. Children mentioned that when a panel is not associated with any exhibit, it should have its own space, rather than just be at a back of another exhibit, and include features that attract visitors to look at it.

Lastly, the *UV* and *IR cameras* share a panel right between them and it is big enough to be in the line of sight of all of the participants. However, Children's focus group members missed it. "I didn't notice a sign for it. And I know it's there, but I didn't notice a sign" (T, C). The author theorizes that maybe, by being a panel for two, it became a panel for none.

7.3.3 Transferable attraction

Sometimes it is possible to use one zone's attracting power to attract visitors to a different, less attractive place. An example in Discovery World was the *Plasma Room* (high attracting power) and the Light Zone (low attracting power).

> Instead of having the entrance [to the Light Zone], that is kind of boring, you could just put a door in the plasma section. So, you can come into this interesting plasma and then you can keep going through into the next room … It'll be a more captivating entrance. (DW, A)

Figure 7-2 explains how this transference of attraction would work by creating access to the Light Zone from the *Plasma Room*. 'Transferable attraction' was not found in the literature.

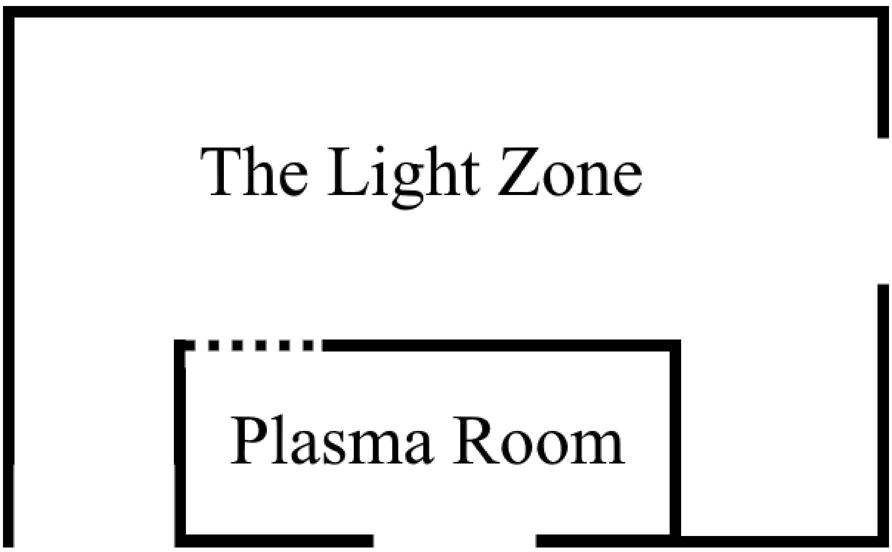


Figure 7-2. General floor plan of the Light Zone and the *Plasma Room*. The dotted line represents a wall that could become an entrance.

7.3.4 Novelty

Humans are curious by nature, always wanting to see something novel (R. S. Miles et al., 1988). The most obvious form of novelty is something never seen before, like *Chicken Embryo*, an exhibit that displays chicken embryos over the first days of development. Novelty is a situational interest trigger (Cardiel & Pattison, 2014), but critical exhibits [like *Chicken Embryo*] can be controversial due to their complex and emotionally charged nature (Pedretti & Navas-Iannini, 2018). “I fundamentally disagree with this killing 24 chickens a week … I think people would get as much out of pressing a button” (SC7). However, visitors engage more with authentic objects, the ‘real thing’ (Gazi, 2014; Schwan et al., 2014), more than with touchable 3D replicas or computer simulations (Lindgren-Streicher & Reich, 2007) or photographic reproductions (Schwan, Bauer, Kampschulte, & Hampp, 2016).

The juxtaposition of real objects with a story gives the idea being communicated a deeper level of significance (McKenna-Cress & Kamien, 2013). “I started off really disliking it [the *Chicken Embryo*] and really struggling with it, but I think it shows us something that you can’t see anywhere else” (SC3).

Despite the variety of responses and points of view critical exhibits can generate, they provide useful context for scientific debates and controversies. They invite visitors to explore beyond prior conceptions (Pedretti & Navas-Iannini, 2018). Actually, Serrell (2006) considers not shying away from controversial issues a point of exhibit quality. “Even when they make claims to scientific objectivity and precision, exhibitions inevitably reflect the beliefs, assumptions and ethical values of the persons making the decisions.” (Gazi, 2014, p. 2). “What we put on display and what we say about it is critical in shaping visitors’ perceptions” (Gazi, 2014, p. 7). A provocative exhibit can become a social object, a platform for visitors to discuss something (N. Simon, 2010). “I really like that [the *Chicken Embryo*], not only because of the biology side of things, but also because of the conversation that we can have around it as to what is ethical” (SC5). The guiding principle in any exhibition should be openness and honesty (Gazi, 2014).

Visitors not only prefer authentic objects, they pay more attention to them and, therefore, learn more (Hampp & Schwan, 2015; Pedretti & Navas-Iannini, 2018; Schwan et al., 2016). “I thought it [the *Chicken Embryo*] would be incredibly controversial. I thought there could be lots of complaints. But generally parents have really enjoyed it and it has been more of a genuinely good learning tool” (SC1).

Novelty can come in more flavours. Sometimes it may suffice to update previous attributes, as seen in this conversation about *Smell This* (T, C).

> ─I didn't like smelling them, but I think that maybe they should add it back. But just with new, gross, smells.
> ─Just with nice smells.
> ─Nice smells for once.

Ideally, exhibits should appeal to different learning styles (McKenna-Cress & Kamien, 2013), and multisensory exhibits are appealing to a broader range of visitor learning styles (S. Allen, 2004). Smell has a particularly direct resonance in long-term memory (McKenna-Cress & Kamien, 2013). Even disgusting smells can be engaging when handled correctly and with humour (McKenna-Cress & Kamien, 2013).

From the very definition of novelty, it is clear that exhibits' freshness cannot last with multiple visits of the same population. However, Science Shows do not necessarily suffer from this issue. They can be adapted continuously and bring experiences that leave a long-lasting impression. "How you can hold flames in your hand without them burning if they are hydrogen gas caught in detergent bubbles (Female, 41+)". Even during the show itself, Science Communicators can adapt the techniques depending on the public. Just before finishing writing this thesis, it was announced that Science Shows would come back to Tūhura, which is a cause for celebration. "It's not a matter of giving people what they ask for, I think it's a matter of giving people what we can (SC5)".

7.3.5 Intergenerational interaction

Parents' education, attitudes, and interest play a large role in children's dispositions and achievement (Alexander et al., 2007; Gunderson & Levine, 2011; OECD, 2013a). As stated in Section 2.3.10, parental presence and involvement with their children's interactions are paramount for learning. Parents often encourage children to persist at exhibit activities and try different ways to interact with them (A. Anderson et al., 2016).

> I think it is a very powerful way of getting young people, young visitors especially, excited about science. And particularly I like the way it [Tūhura] sort of forces intergenerational learning. The parents can't just leave the kids to run around. A lot of the exhibits demand the parents to be there and work with the kids, or the grandparents or carers, whoever they are. (DM3)

Children benefit when parents mediate the exhibit for them; this makes family activities important for children to learn about science (Fenichel & Schweingruber, 2010). Adults, too, can benefit from informal science programs (Fenichel & Schweingruber, 2010). However, they need to feel attracted to the exhibits. "In Discovery World you had the kids doing something and the parents basically just waiting or doing something else" (DM2). The intention for Tūhura was to make the science centre "feel accessible for everybody. So that whether you're 90 or 9 years old, you feel like you can engage in it" (DM5). The goal was achieved in Tūhura, as shown by this comment on the Questions & Answers board: "I love this kind of exhibition! Thank you Having fun! :) Still at 31 years old ♥".

The importance of having an intergenerational space goes well beyond encouraging parents to get involved with their children's learning. Adults deserve a space where they can engage with science, regardless of whether or not they have children. The intention of an exhibition to be for adults as well as children needs to be clearly stated, as adults tend to think these venues are geared toward children (Fenichel & Schweingruber, 2010; Serrell, 2014). The following suggestion was left by two visitors in their 30s, with no children:

> Host adult only evenings at afternoons → where the adults can confidently play with the activities without feeling guilty to take over children's time. Wine + cheese won't go astray. Target adult education theme the topics + build on the educative topics. Look at the Natural History Museum in London. (T, F, NR)

7.3.6 Motivation to read panels

It is widely assumed that visitors don't read the labels (Falk & Dierking, 2016). But actually, visitors read labels if they have the right motivation (Falk & Dierking, 2016; Screven, 1992, 1999). Sometimes it is the lack of incentive, not the scientific information, that's the problem. "The scientific explanations are really good, [but] there isn't much incentive or anything to go read it" (DW, A). Whether a visitor chooses to read the labels or not depends on the benefits (satisfying curiosity, learning something new, giving meaning to the object, being provoked by a question) surpassing the cost (amount of reading, ease of processing information, proximity of the text to the object) (Bitgood, 2016). There should be some sort of reward, so that the visitor feels glad they took the time to read through to the end (Serrell, 2015).

How to present the information is also of paramount importance, especially for children. "They don't exactly just go read a paragraph on a board, normally" (DW, A). Adults and science-engaged visitors may be ok with the usual textbook style of traditional labels, but not visitors in general. "They're good for us, because we wanna know how it works, like what happens, but for little kids, they don't grab them like 'Well, I wanna know [how] this works'" (DW, A).

The vocabulary used in the labels also needs to be familiar (Serrell, 2015). One unfortunate thing that often goes wrong in educational exhibits is they introduce new, obscure topics and then explain them using other unfamiliar terms (R. S. Miles et al., 1988). "Some of them [the labels] use words that are like, some people wouldn't understand. So, they're too scientific, sort of?" (T, C).

To make labels more attractive to children, focus groups suggested presenting the information in more novel ways, like touchscreens, "which kids are very interested by and want to read. Because it's fun to play with, instead of painted on the wall" (T, A). By presenting information with sound and animation, touchscreens would increase the forms of interaction. "That'd give them more reason to go and listen to it. They wouldn't necessarily have to read anything. They could just listen" (DW, A). This is feasible, as visitors can comfortably understand about 150 words per minute (Serrell, 2015). The technology can also allow for the display of multiple languages and provide more detailed information on request (Fenichel & Schweingruber, 2010). However, the advantages may not be worth the expenses when the extra information can be just put on a website (Serrell, 2015).

Well-designed labels can be engaging experiences, but those containing only text have low attracting power (Serrell, 2014). Communication through graphics can reach people who do not rely on words, and add a new dimension to verbal and concrete experiences (Serrell, 2015). However, they shouldn't be used alone; words are still required to help reinforce, clarify, or decode the image (Serrell, 2015). A cheaper alternative to touchscreens could be colourful comic strips. They are cheap to produce and may still be attractive for children to read.

Focus groups also suggested that, in some cases, well-known background historical facts can be used to make a connection. "[*Newton's Prism*] was very cool how they related that to Newton's first experiment. As if you were re–[en]acting that and that can engage, at least with me, with the whole exhibit, a lot further. Because I felt that connection too" (T, A). Investigations into the use of narratives to relate scientific explanations and biographies of renowned scientists has only recently started (Schwan et al., 2014), but it can translate into heightened outcomes (Arya & Maul, 2012). Something similar can be done with other exhibits, like *Vacuum Drop*. "You could just have like a clip of the experiment they did on the Moon. And then that recreates it. To show the differences between [presence and absence of air]" (T, A). A good label fires visitors' imaginations (Serrell, 2015).

7.3.7 Spatial layout of exhibits

Two plasma exhibits were brought over from Discovery World to Tūhura. Interestingly, these exhibits were part of the very popular Plasma Room, but at Tūhura they lost attraction, especially *Plasma Plates* (renamed in Tūhura as *Plasma: The Fourth State*). The reason might be the difference in how Tūhura's spaces were laid out, as the spatial arrangement of exhibits exerts an important influence on attraction power (Peponis, Dalton, Wineman, & Dalton, 2004).

In Discovery World, the *Plasma Room* was darker, making *Plasma Plates* more attractive (It glowed with greater intensity, see section 7.3.1). In Tūhura, *Plasma Plates* not only lost part of that attraction power due to increased lighting; the layout was changed from an enclosed space (the *Plasma Room*) with only four exhibits, to an open space where the exhibits have to compete with others. While variety increases viewing time, this happens only to a point, after which it may have a negative effect (Robinson, 1928). This is probably because too much variety introduces perceptual distraction (Bitgood, 2016).

Falk and Dierking (2016) present four types of floor plans. The 'survey plan' is an open plan with a completely free choice of paths where visitors quickly orient themselves and decide their pathway. Tūhura is not exactly this type, but a good section of it is open. "Open plans pose some risk of museum fatigue in that unobstructed views can yield few surprises, large spatial volume without visual variety can overwhelm visitors with monotony" (McKenna-Cress & Kamien, 2013, p. 151).

Considering that a particular floorplan is better for learning than others is still controversial (Allen, 2004). Still, Tūhura's open space might be offering an excess of choices. Objects presented at the same time compete with each other and cause the decision-making workload to be higher. This may decrease attention to exhibits and discourage visitors from viewing them all (Melton, 1935; Robinson, 1928). Selective choice refers to people becoming more selective because there are too many options, in which case they will choose the ones with higher utility/cost ratio (Bitgood, 2016). Tūhura's *Plasma: The Fourth State* is not strong enough to overcome its competition.

Cognitive workload can decrease if exhibits are divided in sections (as the Light Zone in Discovery World), and therefore complement each other rather than compete. Adolescents in the Tūhura focus group made the following suggestions (T, A):

> —If the whole thing was split into specific zones, then there wouldn't be a single part as more attractive.
>
> —Everything is just mixed. You've got skulls on a bike next the ball that goes above your head that changes light. You're not focusing in one section or science or anything, you're just constantly being distracted by the next thing … Not taking time to understand it.
>
> —So, if you are reading all the things about light at once, you're more likely to take them in. Because they all relate to each other and not just random bits of information. Also, the Light Zone [Unseen Forces Zone] could probably be like a bit a more closed off. Like last time [in Discovery World] you kinda go into like a tunnel that was quite dark.

Non-obvious patterns are less effective at capturing attention than sequential elements because they require visitors to perform a simultaneous search (Bitgood, 2016). "Boundaries between sections will be much clearer to visitors when the architecture, lighting, and colors of the exhibition support the logic of different thematic areas" (Serrell, 2015, p. 78). Simply put by a visitor, "Just have stands separated with respect to the types of activity and area of science" (T, M, 42).

Care should be given to how exhibits relate to others in the surrounding area (Shettel & Reilly, 1966). Exhibits sharing space should be conceptually coherent (Allen, 2004). This was noted in the focus group when talking about the relation of the exhibits as a whole (T,C):

> MC: It doesn't really make any sense.
> C: With the - the- Like, it's something about the Big Bang, and magnetic things, and then Pac-Man.
> C: The Big Bang, Magnetic Things and Pac-Man [said sarcastically, implying Pac-Man does not fit with the scientific things].

Cognitive overload is a huge problem, especially in hands-on science museums (S. Allen, 2004). Visitors tend to explore a great number of different exhibits for a limited time, instead of exploring a few extensively (Serrell, 1998). But, by having well-defined zones, the number of choices decreases; then, there is no information overload or excessive distraction and no need for selective choice (Bitgood, 2016). Having physically separated zones makes it not only easier to concentrate, but it is more engaging.

> It makes it a lot more interesting because you know [where] you are, instead of having just a collection of everything. Being able to focus in on understanding light and all of that and knowing you're in the area leads to just be more engaged. Instead of your concentration and focus just being, like shoved between everything that you see and just running to see everything you want to play with. So, just having a focus on light, instead of just seeing everything and then not concentrating on one thing. (T, A)

7.4 Exhibit Characteristics Related to Holding Power

The second and third stages in the path towards learning are narrowing down attention for a few seconds (focus stage), and then getting deeply involved in the content (engage stage) (Bitgood, 2010, 2016). Holding power matters because it allows the transition from being attracted to an exhibit to spending a longer period of time at it (Serrell, 1998).

7.4.1 Enjoyment

Visiting a science centre is a free-choice experience. Visitors expect outcomes that fulfil their personal desires (Falk & Dierking, 2016). "There're not great museums like [this] in Japan. I want to come here again, and try everything" (T, F, 16). Thus, "exhibits should not only be informative but also provide a pleasurable experience" (Schwan et al., 2014, p. 77).

> Hi, just wanted to say what an excellent centre this is. We lived in Australia (Canberra + Melbourne) and your centre is better than anything we ever visited there. We also [unintelligible] it better than Te Papa! Keep up the good work. We will be back! (Tūhura's Q&A board).

However, expecting a science centre to provide entertainment while at the same time being a respected institution where visitors take away canonical science is part of the constructivist dilemma (S. Allen, 2004). As mentioned by a Science Communicator, "This is where you come to learn and to have fun, *but to learn*" [emphasis added] (SC4). Nonetheless, these two perspectives can be compatible (S. Allen, 2004). "Enjoyment and other affective responses may evoke positive feelings and attitudes that may lead to subsequent, deeper encounters with science" (T. W. Burns et al., 2003, p. 197)

Visitor expectations can be quite varied. One simple source of enjoyment is just having fun. "[The *Monochromatic Room*] is kind of all interesting but at the same time it's really fun" (T, C). Multi-sensory modalities increase attention (Peart, 1984). For example, tactile experiences can be immensely effective at getting visitors to engage with content (McKenna-Cress & Kamien, 2013). "[*Magnetic Sculptures*] is so fluffy and satisfying to touch and it's like interesting how it sticks to your hands" (T, C).

> If physics instruction is mainly aimed at presenting scientific natural laws and reconstructing mathematically, then the majority of students -both boys and girls - lose interest in physics. In contrast, if physics knowledge is taught in such a way that students can recognise a direct connection to practical life situations in which they are personally interested, then there are good chances that their interest will remain stable or even increase. (Krapp & Prenzel, 2011, p. 25)

Thus, at an individual level, personal taste can be source of enjoyment. "[I like] the Light Table [*Ray Table*], but that's because I'm an astronomer" (DM3).

7.4.2 Comfort

One of the biggest enemies of learning at a museum is known as 'museum fatigue'—feeling physically or psychologically tired (Gilman, 1916; Melton, 1935; Robinson, 1928). "Museum fatigue is an important factor that limits the degree to which visitors can effectively learn any form of science" (S. Allen, 2004, p. 20). Both physical and psychological fatigue affect learning (Falk & Dierking, 2016), but in this section we focus on the physical part (that ultimately produces psychological fatigue). See Section 7.3.7 for discussion related to cognitive overload.

Although it is recognized that physical fatigue can affect visitor cognition (Schwan et al., 2014), substantial empirical research is lacking on this topic (Bitgood, 2009). And it is something worth paying attention to at Tūhura, as there were comments in the surveys as well as on Tūhura's Questions and Answers board about things that can be involved in museum fatigue. "More drinking fountains for people to use and more space for clothing and belongings" (T, F, 16). "More toilets outside the tropical forest" (T, F, 50). Seating was especially mentioned in surveys, "Seats are good for grandparents" (T, F, 80). "More comfy chairs for parents. + older people please" (Q&A board).

Physical fatigue decreases attention (Bitgood, 2016), but this can be prevented by providing comfort (McKenna-Cress & Kamien, 2013; Serrell, 2006). A non-participant couple in their 70-80s approached to chat with the author. At some point, their talk diverged from the virtues of Tūhura to mention that Nazis had a very special form of torture: standing up all the time. Regardless how much they praised Tūhura, they felt also tortured.

> There could be a dozen [benches] more. We walk more than 4km a day, but here it feels more like standing all the time… Our attention span is longer [than kids'], but when it wears out, it wears out… Standing gets tiring, really tiring… How boring you find it [the science centre][158]. (T, F, NR)

Comfort is multidimensional, especially in older people; it is not only about sitting. "We need better seating in there [Tūhura]. We need to have better lighting. And we need to reduce the noise levels" (SC7). The importance of comfort factors, such as lightning and seating should not be underestimated (S. Allen, 2004). "Creation of atmosphere by use of color, lighting, staging, or arrangement is heavily used in exhibit design, yet its consequences for learning and knowledge acquisition have been largely ignored by research" (Schwan et al., 2014, p. 77). Exhibition lighting also influences how enjoyable a visit can be (Kottasz, 2006), and non-uniform lighting largely influences visitor's attention (Boyce, 2004).

Reading panels also needs to be a comfortable experience. Typography, vocabulary, and colours should make the text legible (Serrell, 2015). Tūhura tried to make panels more moderate in their use of colour, but they became "a bit hard to read, because of the wood and the white [the background and font colours]" (T, A).

> Regardless of the choice or combination of colors, the most important thing is contrast. Many museums make the mistake, for aesthetic reasons, of having soft-looking labels, such as white type on gray type, or brown type on tan, which renders the labels less legible than they should be for the reader's sake. (Serrell, 2015, pp. 274-275)

[158] The author wrote down her comments and, to his best knowledge, are verbatim.

However, at Tūhura there is not enough contrast between the text and the background. “It’s not just the high or low, it’s also they’re just hard to read … Some of them are on a, like a white-y silver kinda board and they’re silver” (T, C). If a panel is hard to read, the effort or ‘cost’ of reading can surpass the benefits and people will not read it (Bitgood, 2016). Lack of contrast at Tūhura’s panels is especially problematic under the low lighting design.

> I’d change the lighting, and probably the signage a wee bit, make it a little bit easier to see and not quite so dark, the lighting. Cause there’s a few people that don’t see well in the dark and they can’t read a lot of the signs, cause they’re quite small. I, myself have issues with them sometimes, cause my eyes aren’t the best. (SC8).

7.4.3 Interactivity: property of the exhibit

I hear and I forget; I see and I remember; I do and I understand[159]. Interactivity is the ability of an exhibit to respond to visitor actions (S. Allen, 2004). Focus groups praised it as a highly important feature. “Interactivity is what made it [the *Plasma Room*] very interesting” (T, A).

> Everything [at Unseen Forces Zone] was interactive, and everything grabbed and held attention. There wasn’t anything that you’d just would look at and think about. There’s definitely something you can put your hands on and experience for yourself, which then leads to wanting to read the information and wanting to understand what you’re doing, which makes everything more engaging. (T, A)

Interactive elements are more attractive to visitors (A. Anderson et al., 2016; DeWitt & Osborne, 2010; McKenna-Cress & Kamien, 2013; Schwan et al., 2014), promote learning (S. Allen, 2004; Fenichel & Schweingruber, 2010), and make the experience more memorable (Maxwell & Evans, 2002). But interactivity is not necessarily hands-on. The use of other senses adds variety in a science centre, and makes it more interactive as a whole. For example, in *Smell This*, “you are actually using your nose. It’s not a common sense that is evoked in a museum setting” (DM1).

However, it is important to keep in mind not everybody has the same physical capacities. For example, smell can evoke strong memories, but some people can’t smell at all (Serrell, 2015). In this sense, *Sound Bite* is even better, as even people with hearing impairment can use it.

> I never experienced one [*Sound Bite*] before personally and my first experience was that it like gave me a ‘Wow’ moment … Just shows you how brilliant science can be in the communication … It puts a lot of adults outside their comfort zone. They’re having to do that quite bizarre thing in public with biting onto something … I´ve never had someone walk away from that and not be wowed with the experience … For

[159] This proverb is often wrongly attributed to Confucius. Its origin is unclear and the current version might come from a mistranslation or adaptation. However, Confucius, Xunzi and Aristotle did say similar things, all pointing to doing as the epitome of learning.

> people that have hearing aids and they use it, they can´t hear anything if their hearing [aid] is on; if [they] turn it off, they can hear … These people can hear sound whereas otherwise [they] could never. Which is very, very cool. It's a very brilliant tiny thing. (DM1)

Ideally, interactive exhibits should reach an 'active prolonged engagement' in visitors, a 'minds-on' state (S. Allen, 2004). However, as mentioned in the quote above, adults are more reserved than children and may need an invitation to participate (Serrell, 2015).

In Tūhura, there were closely related pairs of exhibits where one has a good level of holding power and the other doesn't. The first example is the pair of plasma exhibits. With *States of Matter*, "you could drag it [the electricity] around more, and the other one [*Plasma: the Fourth State*] was just like being going out from the focused point. So, you can't really do much with them" (T, C). Notice how the difference in holding power is being able to modify the properties of the system, not just activating it.

The second example comes with the *Infrared (IR)* and *Ultraviolet (UV) cameras*. In the surveys, the *IR camera* received 11 mentions that demonstrated learning. For example, "That my body/skin is really warm" (T, M, 11). The *UV camera* was mentioned only once. The *IR camera* "becomes very interesting, because you get to see body temperatures. And then, I don't know, hold your phone to it and see how hot your phone is" (T, A). In contrast, this discussion in the adolescents focus group indicate that it was not clear what the *UV camera* was about (T, A):

> -The *Ultraviolet Camera*. I didn't see how you can see the ultraviolet light. It's just like a black and white camera.
> -Yeah, I think the only thing I picked up on this is I couldn't distinguish between colours and that was about it.
> -I didn't even know that the UV was a UV camera[160].
> -Even after reading it [the panel], I wasn't really sure what was happening. Cause it just looked like…
> -The night vision.
> -Yeah, it just looked like it wasn't even night. So, it just looked like a camera.
> -And it didn't say like what you might be able to use to see the ultraviolet light. The other one's clear and it's like the *Infrared Camera* you can see the heat and stuff. That's quite clear.
> -I thought, even after reading the explanation of all of it, [that that] was what ultraviolet light is and not really shown, it wasn't demonstrated, like demonstrating what the sign was saying, so it just confused me, overall.

Notice how much of the difference between these two exhibits comes from the level of interactivity. The *IR Camera* was clear; it showed differences in temperatures and reacted to visitor's actions. "For example, with my ponytail, I'd hold it up and then it would be purple

[160] Later, after the Focus Group, this participant mentioned he thought it was a surveillance camera.

instead of orange” (T, C). In contrast, the *UV camera* was unclear because it didn’t react to user actions.

7.4.4 Interactivity: the user as part of the exhibit

“We talk about making exhibitions come to life for the visitors; what about the way *visitors* bring the *museum* to life? [emphasis in original]” (Serrell, 2015, p. 66). When a visitor actively participates *in* the exhibit, the experience becomes more memorable. Being directly involved with the phenomena amplifies the opportunity for visitor understanding—physically, emotionally, and intellectually (McKenna-Cress & Kamien, 2013). An example of this type of interactivity is *Sound Bite*, where the visitor’s jaws become part of the exhibit. “You can hear through your bones” (T, M, 9).

Visitors can also expect this kind of participation from an exhibit, like in *Frozen Shadows*, where you can choose to leave an image of yourself with four arms stamped on the wall. But such participation can also be something fortuitous or unexpected, like in the *Plasma Room*, where the black light illumination made visitors’ teeth, white clothes, and other things to glow. “[I learnt] that my [phone] case lights up in the light room [*Plasma Room*]” (DW, M, 13-18). This interactivity can also come from other visitor. “[The *Monochromatic Room*] was super cool. I didn't realise I saw black and white until I saw my mother” (T, F, 11).

7.4.5 Social interaction

Given that humans are social beings, exhibits should encourage and promote social behaviours (McKenna-Cress & Kamien, 2013; Serrell, 2006; N. Simon, 2010). Exhibits that promote social interaction tend to be more engaging (McKenna-Cress & Kamien, 2013; Schwan et al., 2014; N. Simon, 2010). “[When] it’s a game between two people, that’s probably more enjoyable sometimes” (DW, A).

Most people plan their visits primarily as social events, rather than as learning opportunities (Schwan et al., 2014). Either way, learning is more than the accumulation of content knowledge; it is a social process (Fenichel & Schweingruber, 2010). Children and adults reason about issues important to them while interacting with other people (Fenichel & Schweingruber, 2010). To science museum staff, social interaction is one of the main goals of science centres (Shaby et al., 2016).

Competition is a strong element in social interaction (Shaby et al., 2017). *Mind Ball* "is really popular, cause it's competitive" (DW, A). Group visitors often end up teaching concepts to each other (Fenichel & Schweingruber, 2010).

Apart from a design that encourages social interaction, exhibits need to allow more than one user to play with it. Allowing for multiple simultaneous users is a desirable feature (Serrell, 2015) that can facilitate learning (Borun & Dritsas, 1997), but interactivity with multiple simultaneous users can also be disruptive (S. Allen & Gutwill, 2003). "Other people kind of push me off so they can do it, and you have to compete [for the space]. It [the *Ray Table*] is not [big] enough for everyone to play" (DW, C). Not having enough space can be a major issue. It is important to allow for several visitors to cluster around an exhibit and, if possible, allow for more than one to interact simultaneously (McKenna-Cress & Kamien, 2013).

7.4.6 Challenges

Ideal exhibits for learning are driven by curiosity and interest that develops into the 'flow' state where both mind and body are fully involved in the activity (Csikszentmihalyi & Hermanson, 1995). A challenging activity that closesly matches a person's skills can help create that flow (S. Allen, 2004).

Visitors want to be intrigued and challenged (Perry, 2012). Randi Korn & Associates Inc. (2006) reported that exhibits which provide challenge, creativity, and manipulation become favourites. But to trigger situational interest, the level of the challenge needs to be appropriate for the visitor (Cardiel & Pattison, 2014).

Cognitively-challenging phenomena are prefered by students interacting with exhibits (DeWitt & Osborne, 2010). The benefits of challenges go beyond holding a visitor's attention. Contextualized, engaging, and problem-centred activities that challenge individuals to use critical thinking and kinaesthetic abilities enhance learning (Gee, 2012; Lave, 1988; Lave & Wenger, 1991; Mattar, 2018) and knowledge retention (Bruner, 1961).

Focus groups suggested some exhibit challenges that could be incorporated. For example, if the *Ray Table* in Discovery World had targets, visitors could be asked to reach them using different combinations of reflection and diffraction. "Have like little detectors that pick up light. And when the light goes through them, it, like, sound pops up [imitating a triumph trumpet] or something" (DW, C).

Activities that are goal-oriented promote learning as well as feelings of self-determination and control (Fenichel & Schweingruber, 2010). Nevertheless, it is important that the challenges are doable, as self-efficacy and confidence can be increased by experiencing success (Komarraju & Nadler, 2013). “[W]hen challenges and skills are in balance, the activity becomes its own reward” (Csikszentmihalyi et al., 1997, p. 35). However, “there is a fine line between open-ended activities that are challenging but not frustrating, especially for young, inexperienced visitors” (Fenichel & Schweingruber, 2010, p. 48).

> *Floating in Copper* is kinda annoying to figure out how it works … You just mess around with it for a bit, see that it’s not working, like you think it should. And then you’re like ‘Ah, I’m done’. So, it’s just annoying to figure out. (T, A)

People continuously balance the effort (time, energy), the reward (value of the payoff), and the likelihood of success (R. S. Miles et al., 1988). “Quite simply, don’t make visitors work hard for very little payoff” (McKenna-Cress and Kamien, 2013, p. 169).

7.4.7 Testing

The way children build rational structures allows them to clarify some aspects of scientific thought (Piaget, 1968). This ability needs to be fostered. “[I expect] Discovery World 2 to treat kids like smart young adults or smart children, whereas I think [the original] Discovery World doesn’t expect as much from the children” (SC4).If science is about creating theories and testing them, then exhibits should allow that and be open to testing.

> We wanted people to experiment, to kind of work it out. As soon as you work it out, you’re starting to do science, aren’t you? Cause you have a theory about what you think might happen, then you start to play with it and actually see if your theory is correct or not correct. And if it’s not, then you change your thinking to try and figure out positive or negative. (DM5)

The best example is the *Torque Table*. “You can play with it a lot and each time it’s different” (DM2). “It was really fun cause there was so many things to test” (T, C). In this case, it is not about achieving a goal, but challenging your personal knowledge.

> I play a lot of pool and I think I know how pool balls work and react. But then you roll them on that [the *Torque Table*] and they just do something completely alien to what you thought or knew about pool balls. They don’t behave the same. They don’t behave normally. I found that intriguing. So, it’s taking normal things and then not behaving in the normal way. (SC8)

Exhibits that produce multiple outcomes are better for fostering learning (Borun & Dritsas, 1997). Even an exhibit that is popular can be improved by allowing a visitor to test more options. For example, to find out what happens with other monochromatic wavelengths in the *Monochromatic Room*, a focus group participant suggested that visitors could “cycle through different colours. So, like you can have a red, you can have it cycle through the colour spectrum. I feel like that would be pretty neat” (T, A).

Tūhura's *Coloured Shadows* is having a big impact on visitor's learning (see 3.3.13 and 4.2). But while visitors can correctly answer that red, green, and blue (RGB) combine to make white, it is a leap of faith that the same does not happen with red, yellow and blue (RYB). It is common that children learn that RYB are the primary colours at school. New Zealand's Ministry of Education publishes a Junior Journal as part of their instructional series for students. A curriculum level 2 article reads, "Red, blue, and yellow are called the primary colours. All other colours are made from the primary colours." (Wall, 2015, p. 3). This was discussed by the younger children in a focus group (T, C).

> ─And then, for the *Coloured Shadows*. I thought that maybe we can have more colours to play with, instead of just the red, blue and green.
> ─No, because the red, blue and green are the main three colours of spectrum or something.
> ─What about yellow?

The exhibit should allow visitors to be the scientists they are. To include yellow and allow them to test 'What about yellow?' Most exhibits addressing primary colours only include red, green, and blue, e.g., 'The Colour Connection: making colored lights', at the Indianapolis Children's Museum' (Fenichel & Schweingruber, 2010).

Nopparatjamjomras and Chitaree (2009) used an LED mixer with red, green, blue, and yellow to help Year 11 students discover the characteristics of primary colours. They include a question of why yellow is not a primary colour. Unfortunately, they do not report student responses. This would be an interesting study. However, this does not mean that we should keep adding more and more colours. Having too many choices in one exhibit can be overwhelming, limiting learning and enjoyment (Fenichel & Schweingruber, 2010; Serrell, 2006).

7.4.8 Diversity of topics and exhibits

Even though providing multiple ways for learners to engage with scientific phenomena supports learning (Fenichel & Schweingruber, 2010), when exhibition experiences and stimuli are too homogeneous or repetitive in nature, object satiation occurs and interest is lost (Bitgood, 2016). This produces museum fatigue (Gilman, 1916; Melton, 1935; Robinson, 1928). Multiple opportunities for exploration are preferable (DeWitt & Osborne, 2010).

Focus groups expressed the view that *Plasma: The Fourth State* and *States of Matter* are too similar. "Even though, actually, it's different, they are very similar. So, if I had to choose one to go to, I'd choose *States of Matter*" (T, C). It is important to keep in mind that it is not a matter of just adding more, different exhibits. It is important to have "a breadth of stuff that encompasses everybody's kind of interests but then enough depth underneath that to, sort of, entice them to learn more" (DM2).

7.4.9 Linked concepts

Science is multidisciplinary by nature. The exhibits of a science centre should be diverse, but still interconnected—not isolated islands. For example, "The *Infrared Camera* is cool, because it allows a start-off point for conversation about light you cannot see" (SC6). Focus groups suggested labels could be used to link exhibits, or even spaces beyond the science centre, as suggested by one visitor: "I also enjoyed the upstairs geology landscape section; could do more to link them" (T, M, 52).However, maintaining clarity of an exhibit's theme in an open environment is particularly difficult, and techniques to help visitors identify the connections may be effective only under specific designs (S. Allen, 2004).

7.4.10 Lighting

Tūhura, as a whole, is darker than Discovery World. But due to the lack of walls enclosing the light exhibits, the Tūhura Unseen Forces Zone is brighter than the Discover World's Light Zone / *Plasma Room* was. The *Plasma Room* "was segregated into a dark room where all of the plasma parts were set together" (T, A).

For some, Tūhura became be too dark. "I wish we had found a middle ground between those two" (SC3). On the other hand, certain exhibits require an even darker space. "It makes it like stand out more when you are in a dark place, especially since the whole thing's about light" (T, A). During the first focus groups, the researcher asked the participants if removing the walls of the Light Zone would help to have it to be more exposed and have more space for the exhibits. The participants realized that that would come at a higher cost. "The darkness that's in there is really effective, so taking that away would not be good" (DW, A).

The takeaway from this section is that illumination needs to be carefully selected, accounting for each exhibit's own needs instead of trying to find a middle ground for all of them. It doesn't mean that each exhibit needs a different lightning pattern, but they can gathered by groups depending on their needs.

7.4.11 Design and maintenance

Two typical problems of bad exhibits are malfunctions and excessive complexity, which make it difficult for visitors to figure out how they work or how they make sense (Falk & Dierking, 2016). Examples of exhibits too complex to figure out (*Floating in Copper*) or that make no sense (*UV camera*) have already been discussed. The problem discussed in this section is malfunction.

Broken exhibits are not only a waste of space, they affect the site as a whole. "[A broken exhibit] really decreases the popularity of the area. So, it's really depending on the fact that they don't break and that they are well maintained" (DW, A). Because "one of the worst things in going to a science centre is getting in there and half the stuff being out of commission" (SC3).

Exhibits need to be constantly in working condition (Serrell, 2006). The risk of having a non-working exhibit goes beyond simply not being able to interact with it. Control enables engagement (Longnecker, 2016). As discussed more thoroughly in Chapters 1 and 5, confidence makes learning more satisfying and successful. Visitors want to feel in charge of their experiences and feel safe and smart (Perry, 2012). But when the exhibit is broken, "[o]ften, visitors will blame themselves for not being able to understand how to operate an inoperable interactive device" (Serrell, 2014, p. 17). A poorly executed exhibition can make people feel stupid (McKenna-Cress & Kamien, 2013).

> If you have an exhibit, it has to work. I mean, it has to be intuitive, 'cause-otherwise some people do believe it's not working because they don't know how to [make it work]. And then they get the wrong message 'I'm not good about this. Science is not for me, I can't'. (SC5)

An exhibit's working condition is not dichotomous, where something works or it doesn't. For example, The *Monochromatic Room* depends on using flashlights, but they in turn depend on having full batteries, and participants found one of the torch's batteries flat. This is an easily solvable problem as suggested in a focus group.

> The torches are somehow connected to [the] wall … If they're gonna be cabled to the wall, you might as well have them plugged. So, you don't have to worry about them getting batteries and then they can always have access to that light, otherwise the room [*Monochromatic Room*] doesn't work. (T, A)

Getting the best out of an exhibit starts in the design stage, where technical issues might be detected.

> In the *Vacuum Drop*, I wasn't really sure what to do, because the feather was getting stuck. So, I turned it over and it stayed at the top and it was not very engaging. I think there's another object in there, but I didn't really notice that when I was turning around. (T, A)

There are cases where the exhibit is working as expected in technical terms, but the problem comes from the design itself. "The light on the table [*Ray Table*] isn't very bright. So, you can't see the reflections very well" (DW, C). "When it passes through the objects it just gets dimmer and dimmer and dimmer. So, you can't really get the actual effect of what's meant to happen" (DW, C).

Unfortunately, *Light Island*, the exhibit that replaced *Ray Table* at Tūhura, is another example of a problematic working condition that comes from unfortunate design. *Light Island* improved in some senses. It is larger and does not have lateral obstructions or walls, allowing more visitors to interact simultaneously. It now includes white, red, green, and blue in the light sources. There were also shapes that were not present before and some of them were coloured. Still, this change wasn't mentioned anywhere—not in interviews, surveys, or focus groups. What the focus groups had suggested for the *Ray Table* was to either get a stronger laser or make the space darker. Instead *Light Island* is now in a brighter spot and the laser was replaced by light bulbs, whose beams attenuate rather quickly. Even without making them pass through objects they don't even reach the walls of the table.

Sometimes, an exhibit will work as designed, and the phenomena are clearly exposed, but how to play with the phenomena needs a better design or explanation. Tūhura's *Coloured Shadows* provides an opportunity to discuss a problem that may be affecting other science centres. The way to increase/decrease the colour intensity is with three circular dials that turn indefinitely to both sides without being clear which way to turn and when you reached the maximum/minimum. "You weren't sure which way was which" (T, C). "It's like 'Oh, this is a tiny bit more'. But I'm not sure how much I'm actually putting it up for each circle" (T, C). "They need like 'More red / Less red, More green / Less green, More blue / Less blue'" (T, C). Buttons and other interactive devices need to be clearly labelled for what they do, and what the user can expect from using it (Serrell, 2015). Labels of 0% and 100% may help. And there should be physical limits when the user reaches those ends. "You could go up and down with the dials … Like a thermometer, except with like a lever, sort of. So, you can pull up and down to change the amount" (T, C).

Another basic recommendation is to minimize the need for buttons. "Kids are really hard wearing on the buttons. Anything with a button just gets smashed, wears out very quick on a day-to-day basis" (SC8). Visitors can spend so long only pushing buttons before opportunity for learning decreases, due to the less goal-orientated activity (Fenichel & Schweingruber, 2010).

Anything that requires software to run requires special attention and someone with the skills to solve its problems when needed, or there is risk of ending up with no exhibit at all. This happened with *Dancing with Lights,* which was out of commission for five months due to a settings problem.

Sometimes, an exhibit can affect the functioning of others around it. Design needs to account for this possibility. As an example, several exhibits in Tūhura use sand. "The sand gets everywhere, and it's getting into the other exhibits" (DM5). *Magnetic Sculptures* is unavoidably messy; a better design would consider what to do with the sand on the hands, and make it clear and visible (T, C):

> —C1: The towels were quite hidden and hard to see, against the wall. So, some people just dusted their hands off on their jeans, so there's probably a lot [of sand] being lost.
> —C2: I didn't even see the towel.
> —C3: That could be an opportunity, like add a towel.
> —C1: There was!

Focus groups also suggested having a bowl to shake sand off on it. "So, you can actually see a sign saying 'Please wipe sand off here'. So that it's not going everywhere" (T, C).

7.5 Aspects Related to Learning Power

Attracting visitors is only the first step to learning, not the first and last. Fifty years ago, Shettel and Reilly (1966) warned of the dangers of rating an exhibit's effectiveness on only its attracting power: "…making exhibits large, noisy, colorful, and dramatic are techniques that more easily lend themselves to attracting an audience than to communicating ideas or modifying attitudes once the audience gets to the exhibit" (p. 480).

Visitors who engage for a long time are likely to be learning (Bitgood, 2010, 2016; Serrell, 1997). Learning power is the shiny link at the end of the chain. It is only thanks to the previous steps that learning is achieved (S. Allen, 2004).

7.5.1 Understanding what science is

Shaby et al. (2016) found that pedagogical staff of science centres perceive that ideal goals for the centre include understanding what science is, what a scientist does, and how science connects to everyday life. Otago Museum is no different and ideally, visitors come away "being sceptical about things and always researching and checking things" (SC1). That way, they won't be easily hooked by fake news and misinformation. To reach that state, a museum should be "a nice environment that allows people to explore things safely and just become fascinated and become interested" (DM1). In the end, it comes down to people realizing that "being a scientist is about inquiring and have a question and sort of doing tests to find out about it or doing lots of research to find your answer" (SC1). So, science is not only for academics; "it's for everybody and we're doing science constantly in our life, even if we're just cooking" (DM5).

When realising that science is everywhere, visitors ideally leave a science centre with a more analytical eye. "[I learnt about] everything you see every day but may not recognise"

(T, F, 32). “The information is not that important to take away. I think what is important to take away is an inquisitive, and experimental, if you like, ideology around how you approach the world” (DM4).

Great benefits can come if visitors understand what science is, how it is applied, and the importance of the scientific method.

> The most important thing is communicating that the reason [why] science is important is because it provides the best way human beings have of understanding the world around them. Science, in its purest form, isn’t about believing something. It’s actually understanding something on the basis of evidence … A lot of problems that scientists have is that people generally don’t understand what the scientific method is and why it is the best way of understanding the world … If everybody in the world understood science, and understood the reasons that science is such a powerful tool, I think a lot of the silly debates that we have about whether climate change is happening, or whether human induced climate change is happening, or whether we should vaccinate our children, or whether there is pollution, or growing levels of pollution, that would go away. (DM3)

7.5.2 Inspiration and passion

Staff from several museums told Shaby et al. (2016) that some of science centre’s main goals should be motivating people to learn science and fostering excitement, fun, curiosity, and interest. Just understanding what science is and how it works doesn’t guarantee people will side with scientists in polarizing debates (Kahan et al., 2012; Longnecker, 2016).

> The issues we have with climate change is because people are not really feeling for the reality of what’s happening there. They just see it as a sequence of logical thoughts, but they feel for something else. (SC5)

Emotions and feelings are generally seen as biased and inferior to the rational processing of information (Haidt, 2001; Roeser, 2012), but they can be a component within effective science communication (Roeser, 2012). Learning is a multifaceted endeavour that involves positive science-related attitudes and emotions (Fenichel & Schweingruber, 2010). “If you don’t have emotions, if you don’t care for anything, then, that tool of science is nothing either. So, the scientific method cannot go alone, it has to be partner with passion” (SC5). In this sense, exhibits need to go further than just display facts or concepts (Longnecker, 2016). “Dispassionate thought alone could produce unwise actions” (Newton, 2014, p. 84). Exhibits need to ensure visitors feel for science. “Inspir[ing] the sense of curiosity is a key element” (DM3).

Science is often seen as rational and unattached to emotions (Koppman et al., 2015). “Science is based on theories that are tested and they produce unemotional facts” (SC7). This leads to the false idea that science can only advance when emotions are excluded (Barbalet, 2002). In reality, emotions are a key element of creating and engaging with science (Barbalet, 2002; Davies et al., 2019; Koppman et al., 2015; Roeser, 2012).

"You want to be teaching them about science, but you want to do it in a manner that's not like what they get in the classroom. A more sort of exciting manner" (SC1). "[T]he sense of fun and wonder is probably the most important thing to get across. And then, on top of that, an idea of there's a methodology that we can use to increase our knowledge" (DM4). Science is not only intertwined with emotions—emotional expression itself is most salient in scientists at the moment of insight or insight's verification (Koppman et al., 2015).

> We wanted to make people think and have fun and enjoy themselves, and if they walked away with what we called the 'Aha' moment, then brilliant. And the 'Aha' could be smaller or could be big. It could change your life or it could be 'Ah!, now I understand what gravity is'. … [It] could be a really simple thing or it could be life-changing and they go onto become an award-winning scientist. (DM5)

7.5.3 The science in science centre exhibits

Helping visitors to understand what science is and to be inspired by it are great goals; but, at a basic level, science exhibits need to explain scientific phenomena. In an attempt to be engaging, it is possible to end up setting science aside, creating non-science related exhibits. Science centres have reported concern about exhibits being created with an emphasis on fun, rather than instruction (Dicks, 2013). An example of this happened at Discovery World with a Foosball Table. "I don't think it [Foosball Table] is teaching science. I think it's in there, kind of, as a throw away, just to entertain adults, and I think it's not contributing to our mission" (SC3).

Visitors want museums to be playful, but they also want to make sense of their interactions with objects and phenomena (Perry, 2012). Staff consider that one of their science centre's main goals should be cognitive learning (Shaby et al., 2016). If parents are taking their children to a science centre, then they expect science to be an integral part of the site, "and not feel like it is just taking them to the McDonalds playground or another playground. [That] there's actually opportunity for them to read, to learn, to engage" (DM1).

7.5.4 Phenomena exposure

One way to facilitate learning is clear exposure to a phenomenon that contrasts with previous experience (DeWitt & Osborne, 2010). Exhibits shouldn't depend on staff explanations to be understood (Serrell, 2014). "Ideally, I would like a space where it functions by itself. So, we don't necessarily have to be there to explain things to people"

(SC3).Old-style exhibit interactives are more effective than their digital counterparts in foregrounding scientific concepts (Dicks, 2013).

> We intentionally tried to stay away from digital interactives … We wanted people to actually engage with the real thing … When I'm engaging with the black sand or the magnet [*Magnetic Sculpture*], I'm feeling and seeing and touching. That's magnetism, and I can actually see the polar forces starting to happen. But if I was looking at a screen showing me that or a virtual version, it just wouldn't be the same experience. (DM5)

An example of the opposite is *Dancing with Lights*. "It's not something that's even that easy to explain. It's almost like you're doing a sales job for the software. So, it's not about talking about science in that sense. It kind of misses the point there" (SC7).

Sometimes, the problem is not technology per se, but designs that rely upon the idea of a "principal user" and prevent other visitors from interacting at the same time. Tech-based exhibits need to enable simultaneous participation (and encourage social interaction) in the form of collaboration or competition and, preferably, facilitate the co-participation of companions and bystanders (Heath & Vom Lehn, 2008; Witcomb, 2006) - but not to such an extent that social engagement buries the scientific content (Dicks, 2013). Unfortunately, this kind of exhibit is also likely to be mistrusted. *Mind Ball* had the required characteristics, and was so popular among visitors it is one of Tūhura's the most missed exhibits. "That old jedi mind ball kit that would require you to move the ball to the opposition's side via brain activity" (T, M, 15). Still, staff had reasons to mistrust it, making them feel that "it's not really a proper scientific model" (DM3). Because

> I really don't think that it was in any way accurate. I think it was just like a little algorithm that would decide who won each time … I don't think it was reading people's brainwaves, 'cause people would win the game with like the headband half on. (SC6)

Another example of a non-clearly exposed phenomenon is *Vacuum Drop*. It is supposed to show how both a feather and a ball fall under the same acceleration in a vacuum. However, to demonstrate this, both tokens must start falling at the same time once the tube is vertical; it often doesn't happen that way. "Cause you're turning it [the *Vacuum Drop*] over and it [one of the objects] starts going before you turn the whole thing over. It doesn't really make sense" (T, A).

7.5.5 Immediate apprehendability

Immediate apprehendability is a term defined by S. Allen (2004) as "the quality of a stimulus or larger environment such that people introduced to it for the first time will understand its purpose, scope, and properties almost immediately and without conscious effort" (p. 20). It is related to the prior knowledge of the user, but also to how intuitive the exhibit is. Having exhibits that require low effort to be understood is a desirable feature that decreases cognitive overload and museum fatigue during a visit (S. Allen, 2004).

An exhibit's design should enable it to be easily comprehended (Shettel & Reilly, 1966). "[S]ometimes too many interactive features can lead to misunderstandings or cause visitors to feel overwhelmed" (Fenichel & Schweingruber, 2010, p. 43). It is more likely a visitor will stay and engage further if they immediately know how to operate the exhibit, rather if it the operation is unfamiliar or confusing (Shaby et al., 2017). "We cannot allow that the 60 seconds that a visitor spends on average in front of an exhibit, be used to understand how the exhibit works"[161] (Patiño, 2013, p. 148).

"Have the exhibits more educational without having to read the signs" (T, F, 11). That was a goal of Tūhura's exhibits. "The design of the interactives is such that we hope people will be able to come and see how they work without having to read text" (DM3). A clear and short message, displayed in a vivid manner, captures the attention of the visitor easier than when they have to read text before interacting (Alt & Shaw, 1984). Labels and graphics can reinforce what is expected from visitors to do intuitively (Serrell, 2014); still, labels should be based on some prior knowledge and be as short as possible (Serrell, 2015). Connecting to prior knowledge increases relevance for visitors (McKenna-Cress & Kamien, 2013).

One way to achieve immediate apprehendability is through 'user-centred design'. In this design, physical forms and locations invite specific kinds of use, making interaction obvious and simple to use (S. Allen, 2004). For example, *Torque Table* has a big disc and different objects scattered around. The disc automatically starts turning when someone approaches, inviting users to place the objects on top of it. "How could you get that wrong? I mean, how could you look at it and not actually know what to do, it's so intuitive" (DM5).

Nonetheless, it is important to remember that although a lesser amount of text is better; it should provide all the information a user requires. Otherwise, the exhibit may become 'underinterpretive'. "Such exhibits are, therefore, only likely to mean something to a specialist with considerable previous knowledge and a conceptual framework in his head" (R. S. Miles et al., 1988, p. 65).

[161] Translated from Spanish.

7.5.6 Instructive labels

An exhibit at Tūhura that could be considered underinterpretive is the *Blue and Red Buttons*. “It doesn’t really make any sense” (T, C). “What do the *Red & blue buttons* have do with science? Is it time reaction?” (Tūhura’s Q&A board).

Minimizing information does not mean removing it completely; sometimes visitors require background information. “There is nothing more ‘demotivating’ than finding oneself in a situation in which one does not know what is supposed to be happening” (R. S. Miles et al., 1988, p. 33).

What might be totally clear to one visitor, due to their prior knowledge, isn’t guaranteed to be so for another. Visitors expressed their desire for more information for several exhibits in survey comments—“Instructions for the spinning seeds [*Torque Table*] was a little obscure” (T, F, 50); “The torch [in the *Monochromatic Room*] was not very clearly labelled and some people might not realise it could be used to reveal the proper colours” (M, 43); “Perhaps add more instructions for some exhibits as I had to figure some out for myself. Otherwise everything was great!” (T, M, 14). The desire for deeper explanation of the scientific concepts was also stated at the Questions and Answers board. “I would really like to know how the monochromatic room works!?” “How does the egg turn into a chick?” “How does the sand stick to the magnet?” (age 8). “How does the sound travel to you if you are blocking your ears?!?!” (age 10).

How messages on science museum panels are presented directly influences reader comprehension (Miglietta, Pace, & Boero, 2011; Serrell, 2015). And, a clear set of rules also helps in engaging with the activity (S. Allen, 2004). It fosters learning to juxtaposition a learner’s understanding of a natural phenomenon with the formal ideas that explain it (Fenichel & Schweingruber, 2010).

Labels acquire special importance in underinterpretive exhibits. Instructions need to be as concise as possible, but never sacrifice clarity. “[I]lluminating complex issues in an exhibition often requires a significant amount of text” (Pedretti, 2002, p. 24). An example of too-shallow instructions is *Shark Hologram*. The label said “What to do: Look at the hologram of a model mako shark. Walk around to find the best place to view this 3D photo”. Most of the focus group participants, and the author himself, couldn’t find the place until the facilitator took us to the exact spot where it could be seen (From any other position the frame looked empty). The exhibit was not popular, but this might be simply because people thought there was nothing there. A simple sign on the floor signalling where to stand would have made the exhibit visible.

Aside from presenting information, labels should offer provocative interpretation (Serrell, 2015). Provocation becomes indispensable when it comes to interpretation (Ham, 2016). Engaged attention can be prompted by instructions for an activity (Bitgood, 2016; Serrell, 2015).

For example, it was previously stated that the *DNA Slide* is very popular, but visitors don't relate it to science. It doesn't mean the slide is not a good addition to a science centre—actually, it can be critical for increasing Tūhura's overall engagement. "When there is one iconic object or experience that is a focal point, directly connected to or representative of the broader exhibition content, it can be the most memorable experience for visitors" (McKenna-Cress & Kamien, 2013, p. 179). The *DNA Slide* is a focal point within Tūhura's exhibits zone; it only needs to be more connected to the rest of the exhibition. At the time of data collection, it didn't have any panel associated with it[162]. It makes sense, after all. Everybody knows how to use a slide. But, consider the following instructions: "Look at the slide and compare it to the DNA figure on this panel, what do you see?" Adding this short text, in big words, in a position where the shape of the slide is clear[163], could help visitors have their 'Aha' moment. A question may serve as a prompt to visitors' attention, stimulate enquiry and reflection, and promote more open-ended questions and explanations among visitors (Hohenstein & Tran, 2007). However, the best questions are those that visitors ask themselves (Serrell, 2015).

Labels can be much more than instructions. Explanations and arguments are essential components of the scientific discourse that fosters scientific literacy (Krajcik & Sutherland, 2010). Focus groups suggested adding interesting facts to exhibits. For example, the plasma exhibits could include "'This one, you are touching three million [volts]' … 'This one you are touching 20 volts'" (T, C). One interesting fact for *Plasma: the Fourth State*, was mentioned often: "Plasma makes up 99.9% of our universe" (T, M, 17).

At "*Plasma: the 4th State*. It says that it's like 'This interactive may be dangerous'. Which to me, like people may not want to do that" (T, C). The inclusion of interesting facts can also be a good opportunity to convey safety information in a non-threatening, amenable way. The suggestion of mentioning how many volts you are touching can lead into talking about current and why it is safe to touch electricity at the museum, but not from a wall socket.

[162] At the moment of data collection there were no panels. Not long before finishing the thesis, a giant panel with the image of the double helix was placed beside the *DNA Slide*. However, it does not include prompting questions, only technical information, such as the minimum height of users.

[163] Position is of paramount importance (see Section 7.3.2). If the panel is placed arbitrarily, visitors might look at the slide from an angle that doesn't allow a connection with the double-helix, and the label would be useless, as what happened with *Hologram Shark*.

It is also important to consider an explanation's level of difficulty. If the information requires non-basic prior knowledge, children won't get it. If it is too simple, older visitors may not want to read it. In any case, the main thing to remember is that making labels child-friendly does not mean making them childish. It is possible to present only the basics in the panel, as visitors prefer concrete information over abstract ideas (Falk & Dierking, 2016). More lengthy interpreation can be provided elsewhere, like handouts (Serrell, 2015). "It would be nice to have a way that people can get some more science, like delve deeper into the science. That's probably gonna be an app or QR code" (DM2). Serrell (2015) considers QR 'ugly and obtrusive', but she mentions it might depend on how good the information is when the code is used.

Unread information is non-existent information. "I think everyone was determined to [onomatopoeia of hitting the *Red and Blue Buttons* fast], but no one is reading [that] competition is part of the evolution or any of the information on the wall" (T, A). In those cases, it is a matter of increasing the motivation to read the panel (Section 7.3.6) or make the panel more visible (Section 7.3.2).

7.5.7 Learning, not just engagement

A common claim is that children at science centres learn without realizing it, just because they are having so much fun. But this is not necessarily so. "A highly attractive exhibit does not guarantee good science communication" (Serrell, 2014, p. 11). Section 3.16 anecdotally tells how no science learning was detected in children during the science centre's sleepovers, where behaviour was completely unrestricted, and fun was maximized.

As discussed previously, objects can have stronger attraction by nature of their size, placement within an exhibit hall, or isolation from other objects (Melton, 1935; Robinson, 1928). Those objects have larger attracting power and may even have larger holding power, but that doesn't mean they actually have learning power. As an example, Bitgood (2016) describes a beautiful 30-foot DNA slide in an European museum. Unfortunately, "the size and aesthetic characteristics may be the focus of engagement to the detriment of understanding the biological importance of the double helix" (p. 123).

Tūhura has its own example: a big *DNA Slide*. Despite its popularity, it barely appeared in survey comments. And none of the comments mentioned anything related to DNA. Both focus groups and staff suggest the science of DNA is not being conveyed. "I kinda see that, but I think most people would think it's [just] a fun slide" (T, C). For example, the following comment was left at the Questions & Answers board. “What is the slide for? I love science btw”.

> I’ve got a love-hate relationship with it [the *DNA Slide*], because I feel it’s not doing what I hoped it did. It’s a really cool exhibit, but in terms of science communication, I would have wished to make it more like DNA. To make more a point of gravity. Of what happens when you’re riding it. I think kids are enjoying it just like a slide, like you would enjoy a slide in a park (SC5).

7.5.8 The telephone game

An exhibit can be intuitive, science-based, simple, with phenomena exposed, and visitors can claim they learnt science from it. But that doesn’t mean they did, or if they did, it doesn’t mean they learnt what was expected. For example, Dicks (2013) observed and chatted with young children while interacting with a Kugel Ball. No one got the scientific principle. Instead, they played with it as a ‘wishing ball’.

An example from Tūhura is *The Void*. It is the most interesting exhibit in terms of interweaving western science and Mātauranga Māori. “It’s a really amazing interpretation of the Big Bang and the Māori vision of creation and it just puts the two side by side really beautifully” (DM6). One of the Museum decision-making staff described it.

> The story of the Big Bang is the story before we know the world as we know it. And then there’s some of the traditional Māori creation stories, which talk about Rangi[164] and Papa[165], who were held in their embrace of love with their children between them, moving in the darkness. Eventually, the children became so frustrated of living in the darkness they plotted to push their parents apart and then the world as we know it became as we know it. With light and earth and forests, and mountains, etc., who are all the children … So, when we go in here, you see the Big Bang as we know, everything happening all at once with the lights that are in there. But you’ll also, through the soundtrack, start to hear the dang, deng, dah, dank [onomatopoeias of water drops] dark dropping. You feel like you might be in a musty cave. It’s like being between Rangi and Papa, as the children. And then everything comes out. You hear some Māori instruments in there, some very traditional Māori, wind instruments in there. You’ll hear rain. So, you’ll start to kind of hear the creation story coming through on the soundtrack as well (DM5).

It is a beautiful description, and it is definitely a beautiful experience. But neither the Big Bang, nor the Māori story, end up being clear, as described in the adolescent focus group discussion (T, A):

> ─[*The Void*] did have a description before you go into the room. About the relation to the Big Bang … [But] I wasn’t really thinking of the Big Bang. I was more thinking, ‘Woo, flashy lights and mirrors’ [chuckles]. I was definitely entertained, but I don’t know how much science I was taking in from that whole experience.
> ─I didn’t see the Big Bang related to the lights at all.

---

[164] Ranginui, male god, the sky.
[165] Papatūānuku, female god, the earth.

—And there was water effects, which kind of confused me.
—Like, water sounds.
—I thought there was everything in the Big Bang, wasn't there?
—It just seems like almost as if it was part because it looked cool. And then you just have to try and find something related to science somehow, you know [chuckles].
—It kind of feels like you're just trying to take something from the SteamPunk HQ[166] in Oamaru. Cause they've got something like that and they've had it for a couple of years.

The issue was also raised by visitors. "Make *The Void* actually tell you about the Big Bang" (T, F, 11). One problem with *The Void* might be the lack of prior knowledge (see Chapter 1).

> Cause, if you know what it's about, like you know detail, then you know what's happening. But if you're a random person, that's like 'The Big Bang. I don't know', then you don't. It looks like just flashing lights. You have to understand for it to be an amazing experience. (T, C)

Information presented needs to be accurate (Shettel & Reilly, 1966). However, not everything comes down to a message's accuracy. "When interpretation makes us think, our thoughts can generate a lot of sentences in our heads—not just the one an interpreter was trying to generate" (Ham, 2016, p. 117). This was also noticed in the focus groups. "I thought it was really good, because there was so many different possibilities on what it could generally do, like people might have different opinions on what the purpose of it would be" (T, C). Children turned the potential threat of unexpected messages into an opportunity, by allowing them to leave with their own interpretations. "Once they've come out, [they would find] maybe like a table or something on the side so they can write what they think it's about." (T, C). A slight variation from this suggestion would be to ask them to 'Write your own label' (N. Simon, 2010). That would reveal what people think of the experience and, at the same time, make them participants.

The *Monochromatic Room* is another interesting exhibit to analyse in terms of conveying science principles. It is a popular exhibit and surveys indicate visitors consider it science-based and with a clear science message. However, a collateral message may also be getting across. "The *Monochromatic Room*, kind of, shines a bit of light on what it's like for colour blind people" (T, A). Focus groups even considered an opportunity to include a sign about colour blindness. A common misconception of colour blindness is to think it is a condition where people simply don't see colours at all[167]. But, only a very specific type of colour blindness would fit with the current version of *Monochromatic Room*.

[166] SteamPunk HQ is a museum in Oamaru, New Zealand, dedicated to steam punk. They inaugurated their 'Infinity Portal' in 2014. The attraction is very similar to *The Void*.
[167] The researcher didn't correct them during the Focus Group to avoid interference with their opinions.

The risk in having visitors make their own interpretations is that they can get just what was intended, only the gist, completely new meanings, or even wrong ideas (Ham, 2016). How much interpretation is allowed to digress from the expected message is something that must be defined, depending on the goals for the exhibit (Ham, 2016).

7.5.9 Personal relevance

Personal relevance is a situational interest trigger (Cardiel & Pattison, 2014). Even though it is probably not possible for an exhibit to cover examples with personal relevance for all visitors, individual relevance can scale. In other words, taking into account what ‘someone’ wants to receive has the potential to reach what is relevant to ‘the people’ (N. Simon, 2010, 2016). In general, real-life examples can provide an anchor that visitors can relate to. “Water is very strong, it requires a lot of knowledge and understanding how to build a dam” (T, F, 65). Even if a visitor doesn’t grasp the full explanation of a scientific phenomenon, real-life examples may allow them to appreciate them. “People that won’t understand the exact science behind it, can still connect with it” (DW, A).

Let’s take the *UV camera* to exemplify this point. By design, the screen displays the image in black and white— not much difference from a surveillance camera. Even though that in-and-of-itself is unengaging, focus groups suggested a way to make it relevant. “With the *UV camera*, you could have interactivity to do with SunSmart and sun safety” (T, C). SunSmart is a non-profit, health promotion program in New Zealand and Australia that promotes being informed about the dangers of overexposure to UV radiation and how to protect our skin and eyes (Health Promotion Agency, 2019). This topic is highly relevant to this population, as New Zealand and Australia have the highest annual incidence rates (age-standardised) of cutaneous malignant melanoma caused by solar UV radiation (Lucas et al., 2015).

Showing how much UV radiation is blocked by different levels of sunscreen lotion or sunglasses would make the exhibit relevant to people. Nonetheless, for the effect to be more noticeable, the exhibit would require its own source of UV light. “You have to be careful, though, with the ultraviolet one. Make sure you don’t sunburn people” (T, A).

Another option comes from an anecdote where an adult visitor approached the author concerned about why the *UV camera* (and not the *IR camera*) was showing a bright spot coming from his phone, even when the screen was off. The reason is that most smartphones have an IR proximity sensor on the front that is on all the time[168]. This sensor uses a small infrared LED and an IR detector. For simplicity, let's consider the LED to be a blackbody[169]. According to Wien's law, most of the light it emits lies in the infrared (i.e. invisible to the human eye) part of the electromagnetic spectrum. However, it also emits lesser amounts of all possible wavelengths (Planck's law), including teeny-tiny amounts of UV.

The amount of radiation from an IR LED is so small that it doesn't pose a risk[170], but the *UV Camera* at Tūhura is very sensitive. The screens don't show absolute amounts of radiation, but relative amounts. The phone's IR LED glows in the *UV Camera* because its emission of UV light is high compared to other objects in the environment. The *IR camera* doesn't detect the LED because the amount of IR radiation it emits is tiny compared to what's already in the environment. This phenomenon can be used to thrill visitors. In Figure 7-3 the author is in front of the *UV Camera*. The bright spot is not the flash; it is the IR LED that is invisible to the human eye.

[168] This is so it can perform features such as turning off the screen when someone makes a call and preventing the touchscreen from reacting to the user's ear.

[169] "Blackbody" is a physics term for a theoretical object that absorbs all incident electromagnetic radiation (i.e. does not reflect any), and emits a continuum of light according to its temperature. It doesn't mean that the object is black; for example, the Sun is considered a blackbody.

[170] Humans also emit mostly infrared radiation and some ultraviolet light. In fact, we emit more UV radiation than any smartphone's IR LED.

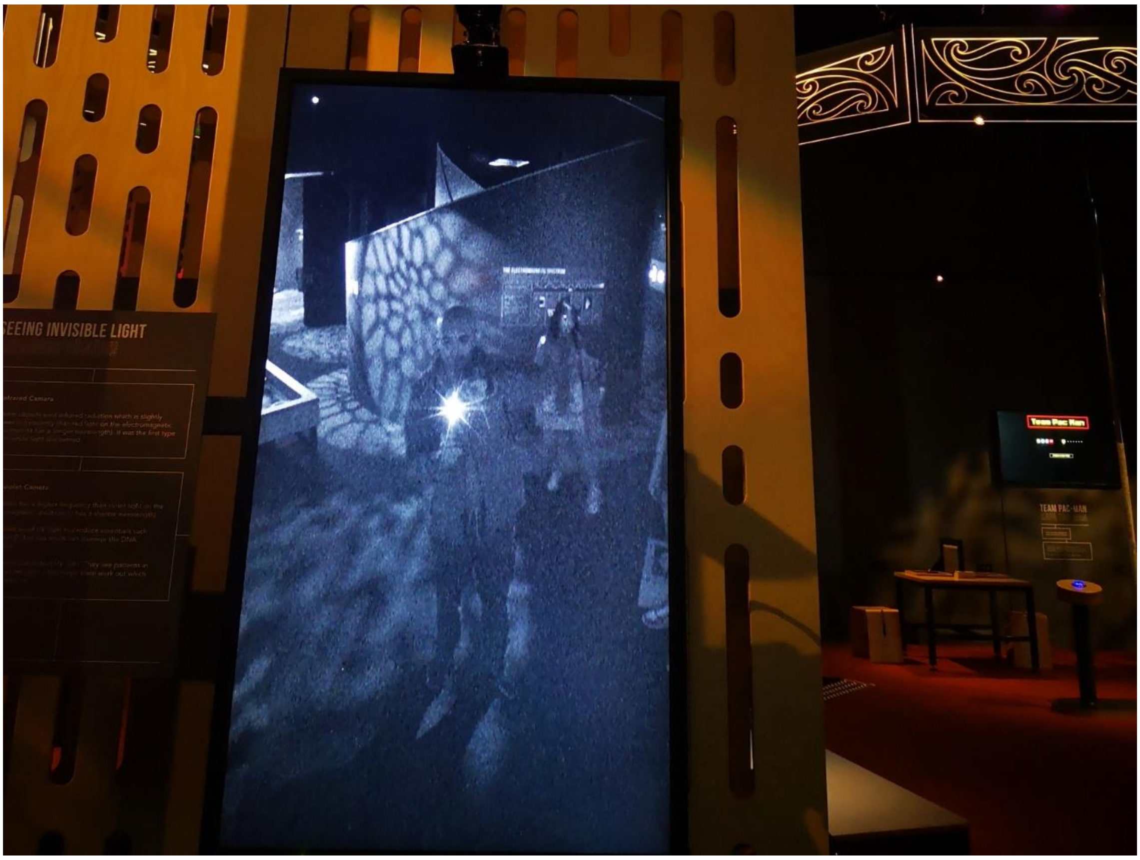


Figure 7-3. Light from the smartphone's proximity sensor at the exhibit *UV Camera*. Without the *UV Camera*, it is not possible to see any light coming from the phone.

Moving the exhibits to a darker spot and adding a third exhibit with a Night Vision camera would help people to fully engage with the exhibits and understand the differences between the wavelengths and the cameras.

7.5.10 Post-visit engagement

Hidi and Renninger (2006) developed a four-phase interest model. The phases are situational interest, maintained situational interest, emerging individual interest, and well-developed individual interest. Short visits to museums trigger excitement, but do not offer enough exposure for longer-term engagement (Fenichel & Schweingruber, 2010).

A visit to a science centre is expected to introduce visitors to the first two phases, but a transition to the third or even fourth phases may require keeping the in-museum experience alive beyond the museum. There are several ways in which it is possible to help the visitor leave maintaining the engagement they felt during the visit.

Museums have traditionally expanded their learning opportunities by lending objects and materials to visitors, such as books or activity kits (Fenichel & Schweingruber, 2010). But, it can also include giving visitors access to special objects in storage facilities (N. Simon, 2010), and the use of internet to post online activities (Fenichel & Schweingruber, 2010) or digitalising objects (N. Simon, 2010). Consider, for example, this note left at the Questions and Answers panel: “Does the museum have a Youtube channel I could watch?”. What matters is not the sharing channel, but that “sharing content helps people learn” (N. Simon, 2010, p. 173).

One option suggested by focus groups is to capture digital memories, especially from artistic exhibits. “I feel like there could be a way to record some videos or take pictures. That’d be very cool” (T, A). And then “e-mail it to yourself and have evidence of doing it later” (DW, A). A video might be too large, but animated gifs are smaller files that can be sent by e-mail. They can also be used to promote the science centre by posting their pictures on Instagram. For example, the Liberty Science Centre has an annual programme, ‘12 Days of Science’, that includes visitors taking pictures of their visit and posting them on Facebook. The favourite three win a prize (Liberty Science Center, 2019).

Another possibility is having short workshops where visitors create their own exhibits, or giving visitors handouts explaining how to build something at home like, for example, a hologram. “[It is] like a square pyramid thing on a screen. And it makes it look like an actual hologram. Like kind of what you’ve seen in the science fiction … You can easily do it at home” (DW, A).

A third option is free souvenirs—something that can be produced easily and cheaply. “I think they should make another exhibit that makes your own UV light glow in the dark slime” (DW, C). Keeping visitors engaged with the museum’s content at home provides a much larger return on investment.

Lastly, a takeaway can be as simple as a word or a challenge. “To engage people with scientific learning outside of these exhibits, suggest a word or phrase relevant to the exhibit that you'd suggest learning [about in] their own time, e.g. how monochromatic light differs from infrared or ultraviolet lights” (T, M, 17).

7.6 Conclusions

This chapter investigated exhibit characteristics in relation to science learning. To better analyse them, they were divided into aspects related to their attracting power, holding power, and learning power. Some of the characteristics align with findings from previous literature, e.g., the use of salient stimulus to attract attention, interactivity to hold it, phenomena exposure to learn, and the importance of readability in labels. However, there are other characteristics that have been discussed minimally in the scholarly literature and deserve further research. Questions for further research include:

- How does distraction hide exhibits in plain sight, affecting attracting power?
- How can attracting power be best transferred to where it is needed?
- How does a particular exhibit's design affect surrounding exhibits?
- How can lateral erroneous messages be detected? (such as an erroneous perception of how colour-blind people see, in the *Monochromatic Room*).

Breaking down what is related to science learning at this science centre was addressed by using SWOT Analysis in focus groups of children and adolescents who were invested in the science centre. This methodology was useful for detecting aspects related to how science exhibits convey science. It would be useful to develop this methodology further to see if it could be implemented with a broader range of participants.

Chapter 1:
INFORMAL LEARNING
OF SCIENCE

Chapter 2:
OTAGO MUSEUM
SCIENCE CENTRES

Chapter 3:
METHODOLOGY

Chapter 4:
SCIENTIFIC KNOWLEDGE:
Are visitors learning?

Chapter 5:
SELF-BELIEFS IN SCIENCE:
To know that you know

Chapter 6:
SCIENCE ENGAGEMENT:
The joy of learning

Chapter 7:
EXHIBITS
AND SCIENCE LEARNING

Chapter 8:
CONCLUDING REMARKS

Chapter 8: CONCLUDING REMARKS

8.1 Key Findings and their Implications

In this chapter, the main findings of the thesis are summarised and their significance is discussed. Findings are related to each of the research questions, leading to recommendations for science centres and for future research.

Scientific literacy is a complex and multi-component construct (National Research Council, 2009). Assessing its changes is no easy task, but to facilitate the enterprise, scientific literacy was divided into three components: scientific knowledge, self-beliefs in science, and science engagement.

8.1.1 Research question one: can scientific literacy of visitors to a science centre be reliably measured?

Scientific knowledge

Learning, traditionally defined in terms of knowledge acquisition (Fender & Crowley, 2007; Illeris, 2018), is one of the main outcomes sought by museums (Jacobsen, 2016). However, assessing knowledge in informal settings has always been challenging.

Multiple-choice questions have been used extensively to asses content knowledge in formal settings (Allum et al., 2008; Ceuppens et al., 2018; Falk & Storksdieck, 2005; Fencl, 2010; Gormally et al., 2012; Hill & Sharma, 2015; Hill et al., 2015; Hill et al., 2014; Kahan et al., 2012; National Academies of Sciences Engineering and Medicine, 2018). Informal learning researchers usually reject them for being disruptive and limited in scope (Fenichel & Schweingruber, 2010; National Research Council, 2009).

Most prior studies with regular science centre visitors have typically relied on indirect measures of learning, such as self-reporting (Dunning et al., 2004; National Research Council, 2009). However, their validity relies on assumptions, such as that a respondent is honest enough and able to provide an accurate self-report (Paulhus & Vazire, 2007). But self-reports can be highly biased, with low-literacy respondents over-reporting (Dunning, 2011; Dunning et al., 2004; Mahmood, 2016; Schlösser et al., 2013), high-literacy respondents under-reporting (Parkman, 2016; Sherman, 2013), or visitors misreporting how familiar they feel with what is being asked (Kelley & Lindsay, 1993; Ladwig et al., 2012; Mbewe et al., 2010; W.-C. Wang et al., 2016).

Scientific knowledge in light and electromagnetism was found to be reliably measured by a multiple-choice test. The instrument was carefully designed to not be overwhelming to visitors, with questions theoretically independent of worldviews. No signs of bias or visitor alienation were found.

Self-beliefs in science

Three self-beliefs were assessed at the science centre. Scientific fluency is a new construct defined as the perceived science-related knowledge that facilitates comprehending natural world phenomena. It is based on the concept of aesthetic fluency (J. K. Smith, 2014; L. F. Smith & Smith, 2006) and it comprises two parts: fluency about scientists and fluency in scientific concepts. The instrument proved to be a reliable way to measure scientific fluency.

Although scientific fluency is considered a self-belief because it is a self-assessment, results from Fluency in scientific knowledge strongly resemble results from the scientific knowledge instrument, suggesting that fluency in scientific concepts may also be a reliable measure of scientific knowledge.

As expected, self-efficacy in science was malleable (Bong & Skaalvik, 2003), but the reliability of the instrument needs improvement because some visitors tended to dichotomize their answers by choosing low-confidence values before the visit (including *I Don't Know* responses), and high-confidence values after the visit. There is no reason to believe that the instrument was not measuring self-efficacy (i.e., it is valid), but reliability may be compromised due to the unexpected dichotomization.

Detecting changes in self-concept is particularly challenging, as this construct is quite stable (Bong & Skaalvik, 2003) - so much so that a Likert-type scale didn't pick up any difference from before to after visiting the science centre. A new format, the Visual Discrete Scale, was developed and tested as an alternative to the Likert-type scale. The Visual Discrete Scale produced similar statistical descriptives, but, unlike the Likert-type scale, it was able to statistically detect a small variation in visitor self-concept in science. This may be due to respondents using the Likert format choosing a 'one-size-fits-all' response after their visit (see Section 5.5). The Visual Discrete Scale didn't show such behaviour. Both scales are valid, but to detect very small changes, only the Visual Discrete Scale may be reliable.

Overall, it can be concluded that self-reports can be a reliable method to collect data about self-beliefs, but the preferred format may vary depending on the construct.

Science engagement

Science engagement is the only construct where measuring it may not be a reliable way to quantify impact of a visit. The chosen science engagement scale had high internal consistency, but the high consistency may be because engagement was already high before the visit. A ceiling effect limits the range of responses and therefore may inflate reliability[171].

A similar thing can be said of the three words to describe the science centre. They were of great help to identify visitor's opinion. But visitors are a self-selected population, and it seems that most of them arrive with an already high level of engagement, producing some sort of ceiling effect[172].

The other reason why these instruments may not have reached their full potential at the Otago Museum is the Tropical Forest. Having the Butterfly House and the science exhibits part of the same experience makes it difficult to separate measures of one from the other. Because the Butterfly House is so popular, the validity of measuring engagement with the exhibits may be somewhat compromised.

What turned out to be a good quantitative measure of engagement was asking visitors before the visit their Intention to do some activities, and their actual Interaction after the visit. That provided a reliable sense of the level of engagement visitors had with specific attractions. The drawback is that the dichotomous nature of the instrument does not allow us to discriminate the levels of engagement for each individual visitor.

Although qualitative data (e.g., open questions) are not quantifiable and therefore not subject to reliability testing, a combination of all methods drew a fairly complete picture about engagement at the science centre.

8.1.2 Research question two: what aspects of scientific literacy can be influenced by visiting a science centre?

Scientific knowledge

As noted above, learning is one of the main goals sought by science centres (Jacobsen, 2016). A multiple-choice questionnaire allowed for assessment of content

[171] To some authors, a ceiling effect is only considered as such when respondents choose the maximum possible value in more than 40% of the cases (Dean, Walker, & Jenkinson, 2018). However, the author suggests that a ceiling effect can be found in responses accumulating near, but not necessarily at, the maximum value. Some authors also claim that the ceiling effect is a threat to validity (Crawford, Briggs, Rodkey, & Steadman, 2007). The author also disagrees with this position. A scale can actually be measuring what it was intended to do, but if the ceiling effect limits the number of possible responses, there is a problem of large enough range for discrimination. In this case, it would be more appropriate to talk about a threat to reliability (J. R. Hayes & Hatch, 1999), rather than validity.

[172] Not in terms of numerical values, as this instrument is not about choosing a discrete response, but about choosing words that denote high engagement.

scientific knowledge objectively before and after visiting a science centre. There was a substantial increase in correct answers after visiting either Discovery World (increase of 15%) or Tūhura (increase of 13%). However, a possible cueing effect may be a factor in this increase.

A new way to visualize learning was devised. The learning flow diagram (see Figure 3.4) is an easy way to not only look at how much correct, wrong, and IDK responses change, but how many answers remain stable versus how many change after the visit. This diagram showed an important flow from wrong and IDK responses to correct answers, especially for visitors who stated they interacted with the exhibits. In these visitors, the flow to the correct answer was larger from IDK (8% in Discovery World, 9% in Tūhura) than from the wrong answers (5% in Discovery World, 6% in Tūhura).

Visitors were also asked to self-report their learning in several ways. A sub-instrument created with items from the Modes of Learning Inventory (MOLI) (J. Griffin et al., 2005) found that 86% of visitors self-reported high or very high learning after the visit.

Although only 36% of visitors reported before their visit that they were coming to the science centre to learn science, 78% agreed after the visit that they learnt something they didn't know before. Visitors also self-reported their learning qualitatively in answers to open questions. Learning at Discovery World was dominated by the Tropical Forest as illustrated by this response: "The butterflies and their habitat that is mostly what we came for on our short visit" (DW, F, 19-40). Whereas the Tropical Forest was still very important in Tūhura, Light and Electromagnetism became the most frequently mentioned topic. "How plasma is a state and when you touch it electrons go through your finger creating a small electrical charge [I] think" (T, F, 16).

Self-beliefs in science

The importance of believing in one's own capacities regarding science has been pointed out in previous research (Christenson et al., 2012; Fenichel & Schweingruber, 2010; Pajares & Schunk, 2001; Schunk & Mullen, 2013), but the effect of visiting a science centre on regular visitors' self-beliefs in science has scarcely been investigated (see Section 1.5). Three self-beliefs in science were assessed in Tūhura: self-concept, self-efficacy and fluency.

Self-efficacy was measured for the confidence in performing specific tasks related to light and electromagnetism. This self-belief is the one that increased the most. Although the large change may partially be due to visitors dichotomizing their confidence in *Can´t* (low confidence) before the visit and *Can* (high confidence) after the visit, it shows how large the impact of visiting a science centre can be.

Self-concept was measured holistically, with respect to science as a whole. People with low perceived ability in science are likely to benefit from an intervention that boosts their academic self-concept (DeBacker & Nelson, 2000). As expected from the literature, self-concept was fairly stable (Bong & Skaalvik, 2003; Lee, 1998) and was barely modified by visiting the science centre. However the Visual Discrete Scale was able to detect a small but statistically significant increase. Given its stability, it can also be expected that the gains obtained in self-concept from the visit will be more difficult to lose later.

Fluency in scientific concepts turned out to be more malleable than self-concept, but less than self-efficacy. Something important to notice is that a LOESS regression on this construct (Figure 5-3) produced a very similar shape to the one produced on scientific knowledge (Figure 4-3). Both have a similar increase after the visit. This suggests that these constructs were correlated; in other words, visitors not only learn science, but they are aware of how much science they learnt. The most noticeable difference between these two constructs is that during adulthood scientific knowledge in light and electromagnetism flattens out, while fluency keeps increasing, although at a lower rate. Importantly, children's fluency in scientific concepts increases after the visit to a level equivalent to someone four years older.

Science engagement

Knowledge alone is not enough for people to feel motivated to change beliefs or practices (Longnecker, 2016); positive emotions are important (Harré, 2011). Change in science engagement was assessed in several ways. As mentioned above, a self-reported engagement scale didn't show any difference after visiting Tūhura. Since free-choice science events have a high rate of enjoyment (Longnecker et al., 2014), it is possible the lack of change is due to a ceiling effect (visitors are self-selected and already highly engaged with science).

The Butterfly House is a very popular attraction that likely has an effect on visitor reported engagement. To break down engagement by attractions, visitors self-reported before the visit their Intention to interact with the exhibits, visit the Tropical Forest or the planetarium, and see the Science Show. After visiting the science centre, they reported what they actually did (Interaction). Visitors tended to do more than they expected to do. Of special relevance is that the number of visitors who actually interacted with the exhibits doubled from those expecting to interact with them. This happened at both science centres, but with an important difference. While prior engagement (Intention) with the exhibits was similar at both science centres: Discovery World (41%), Tūhura (46%), the post-engagement (Interaction) of exhibits at Discovery World (80%) was lower than that of the Tropical Forest (97%). In contrast, Tūhura's Interaction with the exhibits (93%) reached the same level as that of the Tropical Forest (94%).

An interesting method that comprises quantitative and qualitative characteristics is asking visitors to describe the science centre before (expectation) and after the visit (reality) using three words (Longnecker et al., 2014). The words whose frequency decreased the most were *Butterflies*, *Science,* and *Learning*. On the other hand, *Fun*, *Interactive*, and *Awesome* increased the most. Visitors' expectations before the visit sound logical in what they expect from a *science* centre with a *butterfly* house, but they leave with a more vivid and emotional perception of science. Qualitative evidence of science engagement at science centres can also be seen in visitor comments such as this: "It was really fun and a different way of learning. I don't think there is anything to add" (T, M, 10).

Visit time also gave key insights into engagement. Tūhura visitors stay longer (66min) than Discovery World visitors (57min). Their sweep rate indexes (SRI) (Jacobsen, 2016; Randi Korn & Associates Inc., 2006; Serrell, 1997, 2010) were similar, 133 sq.ft/min for Discovery World, and 141 sq.ft/min for Tūhura. Both are excellent when compared to the average of 400 sq.ft/min that Serrell (1997) found for large non-diorama exhibitions.

These SRI include the effect of the Tropical Forest. However, the influence of the exhibits can be seen in that Tūhura visitors who interact with them stay 13 minutes longer than those who don't ($t(742)=3.542$, $p<.001$, $d=0.516$, $d_{CI}=0.144$). Recurrent visitors also stay longer; those on their fourth or more visit to Tūhura stay 11 minutes longer than first-time visitors (Games-Howell test, $p=.008$).

8.1.3 Research question three: what aspects of scientific literacy are influenced by age and gender?

<u>Age</u>

Age is an important variable influencing how much people know (Lindon, 1996) and how they learn (Fenichel & Schweingruber, 2010; Lindon, 1996). This relationship was clear for both scientific knowledge and its related self-concept, scientific fluency; older visitors obtained higher scores. Interestingly, a LOESS regression suggests that this increase with age is heavily influenced by formal education, as both increase rapidly from young age and until twenties. Afterwards, knowledge reaches a plateau and fluency increases at a low rate. This does not mean that adults stop learning; the instruments were merely focused on a confined scope of knowledge (light and electromagnetism). Human knowledge is vast and adults prioritize learning in subjects they are personally interested in (Flynn, 2012), tending to become experts in specific domains (Fenichel & Schweingruber, 2010).

The LOESS regressions showed that, while scientific knowledge and fluency in scientific concepts vary with age, how much visitors learn at the science centre is independent of age. “Raw potential and talent are only a small part of what it takes to become proficient in a skill” (OECD, 2013b, p. 64). In the end, everybody can learn (Ramey-Gassert, 1997).

No evidence was found indicating that self-efficacy in science, self-concept in science, and science engagement varied with visitor age. However, while engagement with science may not depend on the age, engagement with the exhibits does. But it is not a monotonic variation; it varies in the sense of the exhibits having to be appropriate and attractive to each generation. For example, Discovery World was attractive mainly to young visitors, while Tūhura is engaging across generations. “Really enjoy bringing my kids here. They enjoy it but so do I as an adult. Two birds, one stone. And we all learn something new” (T, F, 27).

Gender

A gender gap in scientific knowledge[173] (S. Allen, 1997; Kurtz-Costes et al., 2008; Skaalvik & Skaalvik, 2004) and self-concept in science has been previously reported in the literature (Jansen et al., 2015; Kurtz-Costes et al., 2008).

This thesis’s research found that females scored lower than males in scientific knowledge, self-concept in science, and self-efficacy in science. However, the amount of increase for these three constructs displayed after the visit was not related to gender. Scientific fluency didn’t depend on gender; the reason is not clear.

To analyse the origin of the gap, self-concept and knowledge were studied in terms of both age group and gender. There was no gender difference in Children (8-12 years-old) for either self-concept in science or scientific knowledge. For Adolescents (13-18), a gap in scientific knowledge appears and remains similar for later age groups. Self-concept in science does not present a gap until young adulthood (19-40) and it remains stable for Adults (41+).

[173] Usually referred to as scientific literacy (see Chapter 1).

Some researchers have found that females are more likely to underestimate their skills in science (e.g. Jansen et al., 2015; Reuben, Sapienza, & Zingales, 2014), which could lead to an assumption that the origin of the gap in knowledge is a lower self-concept in science. This would be consistent with confidence in oneself being reciprocally reinforcing with achievement (DeBacker & Nelson, 2000; Huang, 2011; Jansen et al., 2015; Jansen et al., 2014; Marsh & Martin, 2011; Marsh et al., 2012; Wender, 2004; Wilkins, 2004) and self-concept in science playing an important role in explaining the differences in interest in science (Krapp & Prenzel, 2011). However, the results obtained in this research lead to a potentially different order of cause and effect.

An alternative hypothesis is that the gap starts with girls tending to be less interested than boys in physical sciences as they grow up (J. Osborne & Dillon, 2008). The lack of interest would lead them to choosing fewer science and math courses in secondary school (Maple & Stage, 1991). Then, females on average score lower than males in physical sciences knowledge (Jansen et al., 2014) because they haven't studied as much science. Lastly, the reason why females have a lower self-concept than males (Jansen et al., 2015) would be because they assess themselves accordingly to their lesser knowledge.

If this explanation of the gender gap in self-concept and knowledge holds, the gap would originate from a gender difference in interest. It is not that females are 'behind' males in science engagement. But, career choices are not only influenced by confidence and interest in science; relative academic strengths play a factor too (Stoet & Geary, 2018), and girls may have a wider perspective of where they could succeed (Mostafa, 2019). They then may choose something different to science just because it is their choice; they have less interest in 'hard' sciences than boys (M. G. Jones et al., 2000; Krapp & Prenzel, 2011; Labudde et al., 2000; J. Osborne & Dillon, 2008).

To reduce the gender gap in self-concept and knowledge in physical sciences, it may be a matter of fostering girls' interest in physical sciences at an early stage (Leibham, Alexander, & Johnson, 2013). However, being interested is a personal matter; we shouldn't try to 'correct' a gap originated by a personal decision that needs to be respected. Nonetheless, when the decision is a product of social stereotypes (Bian et al., 2017; Kurtz-Costes et al., 2008), lack of opportunities or discrimination (Reilly, 2010; UNICEF, 2007; Vimala, 2010), or low self-esteem in science (Bamberger, 2014), then it is a problem.

8.1.4 Research question four: what characteristics of science exhibits influence visitor learning?

From a series of SWOT focus group conducted with children, interviews with museum staff, and visitor comments, 28 exhibit characteristics that influence learning were identified. Based on the work by Shettel et al. (1968) and Bitgood (2016), the characteristics were divided into three groups, depending on where they had major influence: those related to the exhibit inviting visitors to interact (attracting power), those that keep visitors engaged for long enough to learn (holding power), and those that propitiate science learning (learning power).

Some of these characteristics have been already discussed in detail in the literature, while others have been minimally analysed or not reported before.

One example of a characteristic relating to attraction power is transferability of zone attraction. For example, when an exhibit zone with high attracting power is close to a low attraction zone, some exhibits in the latter zone can go unnoticed, becoming hidden in plain sight. It was also found that this proximity can, in theory, be used to favour the low attraction zone by transference. For example, a walkway from Discovery World's *Plasma Room* (high attraction power) to the rest of the Light Zone (lower attraction power) could have made visitors enter the Light Zone with a transferred feeling of high engagement that they got in the *Plasma Room*. An analogy would be a great stand-up comedian engaging their audience, then leaving them 'in the mood' so the next comedian doesn't struggle to make them start laughing.

A special type of interactivity that's often overlooked is when a visitor becomes part of the exhibit. An example is *Sound Bite*, where visitor's jaws vibrate to produce sound. Visitor's action is not limited to activating the exhibit (e.g. pushing buttons). It is visitors themselves who physically form part of the working exhibit.

In terms of learning power, this research extends the literature by highlighting the possibility of exhibits conveying an unexpected message. For example, the *Monochromatic Room* is highly engaging, and visitors get the idea about the relation between objects' colours and the type of light they are illuminated by. However, it seems that some visitors were misunderstanding that colour-blind people see the world as illuminated by monochromatic light. This is not correct.

Focusing only on whether or not visitors are grasping the idea the exhibit was trying to convey is necessary, but not enough. Detecting collateral messages that are not part of the design may be just as important as conveying the message that was intended.

8.2 Strengths and Limitations

8.2.1 Strengths

Assessment in New Zealand

Museum research in New Zealand is mostly limited to non-science museums (e.g. MacDonald, 2018). There is evidence suggesting that New Zealand students' level of engagement with science depends more on out-of-school activities than those based in the classroom (Woods-McConney et al., 2013), but there is a gap in rigorous research about the effect of the country's science centres. This helps fill the gap and acquires special relevance if we consider that New Zealanders believe science is an important element used to create a sense of national identity (Ministry for Culture and Heritage's Cultural Statistics Programme, 2009). Science centres and museums are expected to become an integral part of the formal science education process in the country (Gluckman, 2011).

Novel methodologies

Assessment of scientific literacy in informal settings is typically conducted through self-reports (Dunning et al., 2004; National Research Council, 2009) and indirect measures that require assumptions to be valid (e.g., visit time being a measure of learning).

Multiple-choice questions have proved their value in objectively assessing content knowledge in formal settings (Allum et al., 2008; Ceuppens et al., 2018; Falk & Storksdieck, 2005; Fencl, 2010; Gormally et al., 2012; Hill & Sharma, 2015; Hill et al., 2015; Hill et al., 2014; Kahan et al., 2012; National Academies of Sciences Engineering and Medicine, 2018). But traditionally, 'formal' testing is not advised in informal settings because of the risk of alienating visitors (Fenichel & Schweingruber, 2010; National Research Council, 2009). This research dared to go against this popular advice and proved that alienation can be overcome with good design. The multiple-choice test can add objectivity to assessing informal learning and be a method to triangulate with other measures.

Novelty in assessment was more extensive than adopting a method from a different space. In this research, the perceived science-related knowledge was successfully measured through scientific fluency - an instrument adapted from the art realm's Aesthetic Fluency (J. K. Smith, 2014; L. F. Smith & Smith, 2006).

Methods were also combined to create new ones. Focus groups allow a group of individuals to discuss, in-depth, a topic in a relaxed and informal setting (Barbour, 2008; Fowler, 2013; Frey, 2018; Hennink, 2013; Hernández et al., 2014; Metcalf et al., 2013; Rio-Roberts, 2011; L. E. Sullivan, 2009). By combining it with a SWOT Analysis, which is used to collect opinions (Jacobsen, 2016), the new SWOT focus group utilises the power of both individual methods.

A completely new format to collect self-reports was developed as an alternative to Likert-type scales. The new Visual Discrete Scale provided promising results that suggested it is potentially more sensitive to small changes in Scientific Self Concept than an equivalent Likert-type scale.

Use of technology

Surveying on paper is simple, but it has disadvantages (King et al., 2013) that can be overcome with electronic devices (Davis et al., 2012; Giduthuri et al., 2014; Leisher, 2014). iPads were used for data collection in this research, which reduced capture errors, was more attractive to visitors, and allowed elements of data collection that would be impossible on paper, like randomizing the order of items in a scale, or recording visit time electronically.

Constructs formality

Self-concept and self-efficacy are closely-related constructs (Bong & Skaalvik, 2003; Jansen et al., 2015). However, they are not the same. Unfortunately, some authors use these terms interchangeably (e.g. Bandura, 1986; Wender, 2004) and mix them in instruments (e.g. Martin et al., 2016). Care was taken in this research to understand the similarities and differences of each construct (not only self-concept and self-efficacy) and to develop valid instruments that measured them.

Pre-beliefs, post-beliefs

The relationship between identity and motivations of visitors has been extensively studied (e.g. Falk & Dierking, 2016). However, few researchers have investigated how visiting a science centre impacts a visitor's self-beliefs in science. Also, many studies focus on school students and field trips and don't consider out-of-school visitors (e.g. Sasson, 2014; Şentürk & Özdemir, 2014). By assessing changes on three different self-beliefs, this research contributes a better understanding of them to the literature.

Care for validity, reliability, and trustworthiness

Validity means that the instrument measures what it intends to measure (Hernández et al., 2014; Neuendorf, 2016). Reliability means that the instrument produces coherent and consistent results (i.e., the results can be reproduced) (Hernández et al., 2014; Neuendorf, 2016). Trustworthiness refers to showing the practices are credible and auditable (Rolfe, 2006). These three concepts are important for a sound research, and all three were attended to.

Statistics rigour

Due to the inherently complicated nature of human studies, there is a lot of misunderstanding in statistics applied to social sciences. For example, Likert-type scales are commonly tested parametrically (Hernández et al., 2014; Moreiera et al., 2019), disregarding that ordinal data are qualitative in nature (Villasís-Keever & Miranda-Novales, 2016). Some authors also exclude the middle point from Likert-type scales (e.g. Cardiel & Pattison, 2014) or suggest removing these responses (e.g. Masuda et al., 2017). This research developed a strict methodology for how to treat data statistically depending on their characteristics (see Chapter 3).

8.2.2 Limitations

Time

The first and more general limitation for this research was time. When the research started, Discovery World was already due to close soon. There was limited time to prepare the design of data collection at Discovery World. For example, several constructs were tested with single items. Even though there was a period of piloting, the lack of time translated into a limited number of responses and precluded the possibility of improving the survey in a second go or complementing it with another one at Discovery World.

Informal participants

Collecting data at a science centre is full of challenges and limitations. It is not only the setting that is informal. Visitors participate freely, and their visit does not follow any rules. Thus, one limitation is the impossibility of having a control group. Moreover, it was not possible to know what changes in scientific literacy came from interacting with the exhibits and what came from other kinds of interaction. For example, the author heard a parent explaining to his children. “You don’t only need an atmosphere [to have an aurora], but also a magnetic field”. But “museum researchers do not always have the resources to conduct rigorously controlled experimental studies” (S. Allen, 2004, p. 25)

Pre-test sensitizing

Matching pre and post reposes in a single group is a widely-accepted experimental design (Creswell, 2009; A. J. Friedman, 2008; Hernández et al., 2014). However, it also implies a real possibility of visitors being cued to what to look for and learn at the exhibition, which could potentially affect the results (A. J. Friedman, 2008). Findings in this thesis need to be taken with the caution of this possibility.

Long-term learning

Long-term learning at school may not come in definite increments in a given lesson, but accrue gradually (Shemwell, Avargil, & Capps, 2015). Long-term learning is influenced by the frequency of review (Yang et al., 2018). Multiple experiences are likely necessary to develop sustained interest (Fenichel & Schweingruber, 2010). However, many visitors may only have one opportunity to visit a science centre. Even if they can, they may not come frequently. This may mean that increases in scientific literacy fade with time. For example, Şentürk and Özdemir (2014) examined the effect of visiting a science centre on attitudes towards science. Self-concept increased from 25.5 points before the visit to 29.1 right after. One week after the visit the change was still significantly higher, but it had decreased to 27.9.

Despite how interesting it would be to know how changes in scientific literacy evolve over time after the visit, assessing retention of information typically requires a follow-up one to four months after the experience (Fenichel & Schweingruber, 2010). This was beyond the scope of this research.

Visual Discrete Scale assessment

Comparing the newly-developed Visual Discrete Scale to a Likert-type scale provided promising results. However, it needs to be acknowledged that the Likert-type scale used a new set of emoji that has not yet been independently tested.

The Visual Discrete Scale is completely visual. Since the total absence of text anchors can be cognitively more complex for respondents used to being given text-based instructions (Friborg et al., 2006), children tend to respond better to scales where each point is labelled, rather than only at the extremes (Borgers, Hox, & Sikkel, 2003). The Visual Discrete Scale may have been more difficult for the younger children in this study, especially aged eight to ten, as from the age of 11 children improve significantly in their capacity to handle complex questionnaires (J. Scott, Brynin, & Smith, 1995). It is acknowledged that not detecting issues with this young demographic may have been due to parental assistance.

Scientific literacy reach

Scientific literacy is such a complex and extensive construct that it would be impossible to cover all of it in a single study. As an example, most visitors get that white light can be obtained from combining Red, Green, and Blue (see Chapter 4), but *Coloured Shadows* is one of the exhibits that produced more questions, like these from the Questions and Answers board: “I want to learn are black and white colours?” (age 8). “How do all the c[o]lours mix together to make white light?”. “How do the shadows get colour to them?”.

Coloured Shadows is an exhibit deriving from a patent by Tsuchihashi et al. (1978). The exhibit is simple to use, but the underlying concept is completely counterintuitive to children, as can be seen in the works by S. Allen (1997) and Feher and Meyer (1992). S. Allen (1997) studied the effect of this exhibit by interviewing visitors to assess visitor’s understanding of the exhibit. This assessment only occurred after the visit. None of the visitors between 7 and 12 years old (n=52) got it right. Those from 13 to 15 years old (n=28) got it right less than 10% of the time, while adolescents from 16 to 18 years old (n=28) and adults over 19 years old (n=280) averaged approximately 40%. Prior understanding turned out to be a large factor in explaining who got the questions right.

Notice that the scientific knowledge instrument of this research was designed to measure content knowledge. It does not measure, and does not intend to measure, conceptual knowledge.

8.3 Recommendations to Science Centres

After the Children’s focus group about Discovery World, one child asked the author if the museum was going to implement their suggestions in Tūhura. At that moment, the redevelopment construction was about to begin and Tūhura’s exhibits and design were already chosen. And so, it was not feasible to implement their suggestions in the short term. However, after Tūhura opened, several things have changed in agreement with this thesis’ findings. For example, the DNA slide now features a big panel with a drawing of a double helix to help visitors realize the relation of the slide’s shape with DNA. Also, an astronaut cut-out is now placed at a doorway that connects Tūhura to the Beautiful Science gallery and the planetarium, making these areas more integrated.

The overall recommendation for any science centre is to listen to its visitors; "audience-centric design processes start by mapping out audiences of interest and brainstorming the experiences, information, and strategies that will resonate most with them (N. Simon, 2010, p. 35). Evaluation should be done regularly, giving a voice to visitors to find out what's working for them and what is not. An easy way to collect data regularly is through floor staff. They are the ones that are in direct contact with visitors. Their points of view are invaluable, and should be paid attention to.

The literature can also be an excellent source of advice. For example, it was mentioned in Chapter 7 that Discovery World's *Ray Table* had characteristics that favoured science learning more than those present at Tūhura's *Light Island*. Both exhibits were similar in concept and it doesn't mean that the former is inherently better than the latter, it is just a matter of improving the design. The *Light Island* exhibit is also present at the Exploratorium, in San Francisco. Serrell (2015) mentions that the original design of this exhibit didn't include labels, but six were added afterwards: *Play with the beams of light; Can you separate white light into colors?; Use the reflectors to bounce a light beam all the way around the central light source, How many times can you bounce a light beam?; Try to 'draw' the letter W with light rays, What other shapes or letters can you make with light?; Try mixing the colors. Can you make yellow light?; Can you get the light beam to focus—that is, come together to a point?* Notice how these labels agree with what was exposed in Chapter 7, especially about challenges, phenomena testing, instructions, and provocative questions. Adding the same or similar labels to Tūhura's *Light Island* could improve the exhibit. Nonetheless, the light sources would also need to be changed. As was mentioned in Chapter 7, the light doesn't even reach the walls of the table, while the image of the exhibit in Serrell (2015) shows much stronger beams (p. 194).

In general terms, exhibits should be designed to be attractive to all generations, promote social interaction, be challenging and interactive, and be open to test different variables. Science phenomena should be clearly exposed, so that visitors can learn science intuitively. However, panels can add valuable and engaging information if correctly designed.

8.4 Future Work

This research highlights multiple avenues for future work. Some are summarized below.

Scientific fluency

Even though concept knowledge and self-reported knowledge are conceptually different (Ladwig et al., 2012), evidence suggests that Fluency in Scientific Concept may be highly related to scientific knowledge. If so, the new construct of scientific fluency may be a viable alternative to reliably measure knowledge in informal settings by self-reporting.

Research on knowledge and fluency in areas other than light and electromagnetism may help confirm if these two constructs are correlated. LOESS regressions could also help detect what knowledge is more heavily influenced by formal education during school years and what by informal education in later stages of life.

An extended version of scientific fluency, administered at different venues in several countries, could also help shed light on how knowledge varies across the globe, what the origin of the 'bump' around 19-25 years is, and why no gender gap was found with this construct.

If further developed, the instrument for scientific fluency can be improved in several ways. The response options could be adapted for young children and include a visual component, as suggested for aesthetic fluency by J. K. Smith (2014). A good starting point would be an adaptation of the concept of visual science literacy (Bucchi & Saracino, 2016). Given that museums are cultural institutions (N. Simon, 2010), culture could become part of the instrument, which would be of special help to understand a bicultural science centre such as Tūhura.

Since perceptual and sensory scientific experiences are driven by 'aesthetic' curiosity (Feher, 1990; Zubrowski, 1982), it seems feasible that scientific fluency has the potential to be further developed as a model of science appreciation, in analogy to art appreciation models (e.g. Housen, 2007; Parsons, 1987; J. K. Smith, 2014).

Visual Discrete Scale and Likert-type scales

The new Visual Discrete Scale outperformed a Likert-type scale in detecting a small change in visitors' self-concept in science. This is worthy of more research. But it is not only the Visual Discrete Scale that requires further research. Reports of the one-size-fits-all behaviour and the Sponge effect in the Likert-type scale were not found in the literature. That is not completely surprising, considering that the effect would be noticeable only in very stable constructs—the kind of constructs that are seldom analysed in pre-test and post-test formats. However, the origin of the behaviour and the effect, and how common they are, deserve a closer look.

New icons for Likert-type scales

Smiley faces in Likert-type scales help children to interpret the scale (Hall et al., 2016; Read et al., 2002; Reynolds-Keefer et al., 2009). However, which Emoji are used can affect how children rate it (Chambers & Craig, 1998). Since Emoji add emotional tone to text (Danesi, 2016), using angry or sad faces on the disagreement side of a Likert-type scale may not be appropriate. A new set of smiley faces was created in this research that is expected to be more appropriate and outperform the traditional set. However, the set still needs to be validated by testing its performance more broadly.

Gender gap and science engagement

It was proposed that the gender gap found in scientific knowledge and self-concept in science is not due to unjustified low confidence; rather, it could originate from a lower level of engagement with the area of science studied in this thesis (light and electromagnetism) relative to other areas (see Section 6.4). Some evidence backed this hypothesis, but it is circumstantial and based on some assumptions, e.g., the gap's origin is different for different generations, visitors that read science centre panels tend to be more engaged with science beyond the science centre[174], or that more science engagement leads to more scientific knowledge. Further research is needed to test these assumptions, which would support or refute the hypothesis.

Engagement does not necessarily equal learning

Some authors suggest that enjoyment and learning are precursor and benefit, respectively (e.g. Colliver & Fleer, 2016; Grinell, 1988; Lillard et al., 2013; Rogers, 2013; Wood & Attfield, 2005), and go so far as to suggest there is no valid distinction between playing and learning in children (e.g. Wellington, 1990). However, the results of this research point more towards that, although engagement helps with learning (Christenson et al., 2012; Schunk & Mullen, 2013), it may not be a sufficient condition for learning to happen (Archer et al., 2016).

It has also been proposed that visit time is not only a sign of engagement, but of learning (S. Allen, 1997). How visit time and learning correlate was tested in this research.

[174] They stayed 6 minutes longer than non-readers and showed a higher level of prior Scientific Knowledge.

No statistically significant correlation was found, but the p-value didn’t completely rule out the possibility. The inconclusiveness may come from the fact that time spent in the Tropical Forest cannot be distinguished from that spent at the exhibits. Reproducing this research in a setting where visit time can be clearly associated only with exhibits of interest could help differentiate whether visit time can be considered a sign of learning.

The hypothesis proposed by the author, and which still needs to be tested, is that engagement is a sign of learning only when people are engaged with the particular topic and is only to the time they engaged specifically with it. Using the previous example, the time at the Tropical Forest can only account for learning about the Tropical Forest, and the time at the exhibits can only account for learning about the science shown at the specific exhibits they interacted with.

The second condition of the hypothesis states that people must not simply be engaged, but engaged *with science* in order to expect science learning. For example, children can be completely engaged with the DNA Slide and spend a long time there, but if they are not engaged with ‘the science of DNA’, then their engagement is not a sign of learning science. However, even if they are not engaged with science, external guidance (like, from parents) or internal guidance (self-control and awareness of the possibility to learn) may push visitors along the track of science engagement and learning.

Sections 3.3 and 4.3 discussed circumstantial evidence with sleepover visitors supporting this hypothesis.

8.5 Final thoughts

“If I have seen further, it is by standing upon the shoulders of giants”, said Sir Isaac Newton. Science communication, like Newton’s science, builds upon its predecessors, on what worked and what didn’t. This research had the good fortune of being conducted before and after a major redevelopment of a significant science centre. Discovery World, the pre-redeveloped science centre, had been around for more than two decades and was showing signs of its age. This venue was also seen by many as a child-only zone, almost like a playground. Adult attendance was mainly related to either bringing children to the exhibits or to visit the Tropical Forest, an enclosed space full of live butterflies.

Tūhura, the redeveloped science centre, kept the Tropical Forest, but unlike Discovery World, the new science centre offers more than butterflies to older visitors. Tūhura visitors increase their scientific knowledge in light and electromagnetism in similar levels to Discovery World visitors. The big redevelopment challenge to see older Tūhura visitors engage with the exhibits was a big success, and it was achieved without undermining young visitors' engagement.

The combination of data from both science centres allowed for an objective confirmation, using techniques beyond self-reports, that visitors learn science when visiting a science centre. Although prior knowledge is related to age and gender, how much visitors learn at the centre is completely independent of these two variables.

Self-efficacy in science was confirmed to be malleable and self-concept in science to be stable. The stability of self-concept in science was used to compare a Likert-type scale format to a new one, the Visual Discrete Scale. The new scale provided similar trends as the Likert-type scale, proving itself to be a good alternative. But it also showed evidence of being potentially more sensitive to small changes, as it was able to detect a small pre-post difference in self-concept in science that the Likert-type scale was not able to.

A new construct, scientific fluency, showed signs of being a better reflection of scientific knowledge than self-efficacy in science. By further developing it, it is expected that scientific fluency may be a reliable way to self-report scientific knowledge.

References

Achiam, M. (2019). Reflections on empowering youth in science museums. In H. McLaughlin & J. Diamond (Eds.), *Science museums in transition. Unheard voices* (pp. 44-48). London, UK: Routledge.

Adams, M., Falk, J. H., & Dierking, L. D. (2003). Things Change: Museums, Learning, and Research. In M. Xanthoudaki, L. Tickle, & V. Sekules (Eds.), *Researching Visual Arts Education in Museums and Galleries* (pp. 15-32). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Adda, G., Adda-Decker, M., Gauvain, J.-L., & Lamel, L. (1997). *Text normalization and speech recognition in French.* Paper presented at the Fifth European Conference on Speech Communication and Technology, Rhodes, Greece.

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point likert-type scale. *Educational and Psychological Measurement, 70*, 796-807. doi:10.1177/0013164410366694

Adigüzel, F., & Wedel, M. (2008). Split questionnaire design for massive surveys. *Journal of Marketing Research, 45*(5), 608-617.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270-301.

Ahteensuu, M. (2012). Assumptions of the deficit model type of thinking: Ignorance, attitudes, and science communication in the debate on genetic engineering in agriculture. *Journal of Agricultural and Environmental Ethics, 25*(3), 295-313. doi:10.1007/s10806-011-9311-9

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*, 4-12. doi:10.1016/J.CEDPSYCH.2010.08.001

Ainsworth, H. L., & Eaton, S. E. (2010). *Formal, Non-Formal and Informal Learning in the Sciences*. Retrieved from http://files.eric.ed.gov/fulltext/ED511414.pdf

Akarsu, B., & Kariper, A. (2013). Upper Secondary School Pupils' Attitudes towards Natural Science. *European Journal of Physics Education, 4*(1), 78-87.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180.

Alexandrov, A. (2010). Characteristics of Single-Item Measures in Likert Scale Format. *The Electronic Journal of Business Research Method, 8*, 1-12.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of educational psychology*, *103*(1), 1-18.

Allen, P., Bennett, K., & Heritage, B. (2018). *SPSS Statistics: A Practical Guide* (4th ed.). Melbourne: Cengage Learning Australia.

Allen, S. (1997). Using scientific inquiry activies in exhibit explanations. *Science Education, 81*, 715-734.

Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education, 88*(S1), S17-S33.

Allen, S., & Gutwill, J. (2003). *Assessing techniques for facilitating meaning-making by science museum visitors.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Allen, S., & Gutwill, J. P. (2009). Creating a program to deepen family inquiry at interactive science exhibits. *Curator: The museum journal, 52*(3), 289-306. doi:10.1111/j.2151-6952.2009.tb00352.x

Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: A meta-analysis. *Public Understanding of Science, 17*(1), 35-54. doi:10.1177/0963662506070159

Alt, M. B., & Shaw, K. (1984). Characteristics of ideal museum exhibits. *British Journal of Psychology, 75*(1), 25-36.

Altman, D., Machin, D., Bryant, T., & Gardner, M. (Eds.). (2000). *Statistics with Confidence: Confidence intervals and statistical guidelines* (2nd ed.) Bristol, UK: BMJ Books.

American Association for the Advancement of Science. (1994). *Benchmarks for Science Literacy*. New York, NY: Oxford University Press.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed., corr. rev. ed.). Washington, DC: American Psychological Association.

Anderson, A., Bequette, M., Cosbey, A., Haupt, G., & Hughes, C. (2016). Create.Connect Summative Evaluation Report. St. Paul, MN: Science Museum of Minnesota.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Andert, D. (2011). Alternating leadership as a proactive organizational intervention: Addressing the needs of the Baby Boomers, Generation Xers and Millennials. *Journal of Leadership, Accountability, and Ethics, 8*(4), 67-83.

Appelbaum, P., & Clark, S. (2001). Science! Fun? A critical analysis of design/content/evaluation. *Journal of Curriculum Studies, 33*(5), 583-600.

Archer, L., Dawson, E., Seakins, A., & Wong, B. (2016). Disorientating, fun or meaningful? Disadvantaged families' experiences of a science museum visit. *Cultural Studies of Science Education, 11*(4), 917-939.

Archibald, T. (2015). "They Just Know": The epistemological politics of "evidence-based" non-formal education. *Evaluation and program planning, 48*, 137-148.

Arts Council England. (2019). *Defining learning*. Retrieved from https://www.artscouncil.org.uk/defining-learning

Arya, D. J., & Maul, A. (2012). The role of the scientific discovery narrative in middle school science education: An experimental study. *Journal of Educational Psychology, 104*(4), 1022-1032.

Association of Science-Technology Centers. (2017). *Science Center Statistics*. Retrieved from ASTC website: https://www.astc.org/wp-content/uploads/2017/09/ASTC_SCStats-2016.pdf

Atari, M., Afhami, R., & Mohammadi-Zarghan, S. (2018). Exploring aesthetic fluency: The roles of personality, nature relatedness, and art activities. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi:10.1037/aca0000200

Bächtold, M. (2013). What do students "construct" according to constructivism in science education?. *Research in science education, 43*(6), 2477-2496.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice* (Rev. ed.). East Sussex, UK: Psychology Press.

Baeza, A. H., Tort, B. E., Romá, V. G., & Benito, J. G. (2001). Escalas de respuesta tipo Likert:¿ es relevante la alternativa "indiferente"? [Likert-type response scales: Is the relevant alternative "indifferent"?]. *Metodología de* encuesta [Survey Methodologies], *3*(2), 135-150.

Ballouard, J.-M., Provost, G., Barré, D., & Bonnet, X. (2012). Influence of a field trip on the attitude of schoolchildren toward unpopular organisms: an experience with snakes. *Journal of Herpetology, 46*(3), 423-429.

Baltes, P. B., & Nesselroade, J. R. (1984). Paradigm lost and paradigm regained: Critique of Dannefer's portrayal of life-span developmental psychology. *American Sociological Review, 49*(6), 841-847.

Bamberger, Y. M. (2014). Encouraging girls into science and technology with feminine role model: Does this work? *Journal of Science Education and Technology, 23*(4), 549-561.

Bandelli, A., & Konijn, E. A. (2013). Science centers and public participation: Methods, strategies, and barriers. *Science Communication, 35*(4), 419-448.

Bandura, A. (1986). *Social Foundations of Thought & Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87.

Barbalet, J. (2002). Science and emotions. *The Sociological Review, 50*(S2), 132-150.

Barbour, R. (2008). *Doing focus groups*. Los Angeles, CA: SAGE Publications.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology, 51*(6), 1173.

Barriault, C., & Pearson, D. (2010). Assessing exhibits for learning in science centers: A practical tool. *Visitor Studies, 13*(1), 90-106.

Barron, A. B., Hebets, E. A., Cleland, T. A., Fitzpatrick, C. L., Hauber, M. E., & Stevens, J. R. (2015). Embracing multiple definitions of learning. *Trends in neurosciences, 38*(7), 405-407.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human relations, 61*(8), 1139-1160.

Baskar, S., Arockiam, L., & Charles, S. (2013). A systematic approach on data pre-processing in data mining. *Compusoft, 2*(11), 335-339.

Battaglia, M. P. (2008). Nonprobability sampling. *Encyclopedia of survey research methods, 1*, 523-526.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143-156.

Belke, B., Leder, H., Strobach, T., & Carbon, C.-C. (2010). Cognitive fluency: High-level processing dynamics in art appreciation. *Psychology of Aesthetics, Creativity, and the Arts, 4*(4), 214.

Bequette, M., Svarovsky, G., & Ellenbogen, K. (2011). *Year 5 Summative Evaluation of Exhibits and Programs.* Boston, MA: NISE Network.

Bergkvist, L. (2014). Appropriate use of single-item measures is here to stay. *Marketing Letters, 26*, 245-255. doi:10.1007/s11002-014-9325-y

Berryman, S. E. (1983). *Who Will Do Science? Trends, and their Causes, in Minority and Female Representation Among Holders of Advanced Degrees in Science and Mathematics.* New York, NY: The Rockefeller Foundation.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389-391.

Bitgood, S. (2009). When is "museum fatigue" not fatigue? *Curator: The museum journal, 52*(2), 193-202.

Bitgood, S. (2010). *An Attention-Value Model of Museum Visitors.* Retrieved from CAISE Informal Science website: https://www.informalscience.org/attention-value-model-museum-visitors

Bitgood, S. (2016). *Attention and Value: Keys to Understanding Museum Visitors*. London, UK: Routledge.

Bitgood, S., Patterson, D., & Benefield, A. (1988). Exhibit design and visitor behavior: Empirical relationships. *Environment and Behavior, 20*(4), 474-491.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science. *APS Observer, 19*, 3.

Bland, M. (2015). *An introduction to medical statistics*. Oxford, UK: Oxford University Press.
Boekaerts, M. (2010). The crucial role of motivation and emotion in classroom learning. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The Nature of Learning: Using Research to Inspire Practice* (pp. 91-111). Paris, France: OECD Publishing.
Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational psychology review, 15*(1), 1-40.
Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension, 50*(2), 1-5.
Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons.
Borgers, N., De Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality 1. *Bulletin of Sociological Methodology, 66*(1), 60-75.
Borgers, N., Hox, J., & Sikkel, D. (2003). Response quality in survey research with children and adolescents: the effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research, 15*(1), 83-94.
Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and quantity, 38*, 17-33. doi:10.1023/B:QUQU.0000013236.29205.a6
Borun, M. (1998). *Family learning in museums: The PISEC perspective*. Philadelphia, PA: PISEC.
Borun, M., & Dritsas, J. (1997). Developing family-friendly exhibits. *Curator: The museum journal, 40*(3), 178-196.
Bourke, S., & Frampton, J. (1992). *Assessing the quality of school life: Some technical considerations.* Paper presented at the annual Conference of the Australian Association for Research in Education, Deakin University, Geelong, Australia.
Boyce, P. R. (2004). Lighting research for interiors: the beginning of the end or the end of the beginning. *Lighting Research & Technology, 36*(4), 283-293.
Boyden, J., & Ennew, J. (Eds.). (1997). *Children in Focus: A Manual for Participatory Research with Children*. Stockholm, Sweden: Rädda Barnen.
Boylan, P. J. (Ed.). (2004). *Running a museum: a practical handbook.* International Council of Museums, ICOM.
Bozdoğan, A. E., & Yalçın, N. (2009). Determining the Influence of a Science Exhibition Center Training Program on Elementary Pupils' Interest and Achievement in Science. *Eurasia Journal of Mathematics, Science & Technology Education, 5*(1), 27-34.
Brady, A.-M. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice, 5*(4), 238-242.
Braund, M. (2012). *Learning Science Outside the Classroom* (1st ed.). London, UK: Taylor and Francis.
Brennan, R. L., & Lee, W.-C. (2018). True Score. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1724-1728). Thousand Oaks, CA: SAGE Publications.
Brenner, P. S., & DeLamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly, 79*(4), 333-354.
Brody, S., & Diakopoulos, N. (2011). *Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs.* Paper presented at the Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK.

Broek, S., Buiskool, B., Van Oploo, M., & De Visser, S. (2012). *State of play of the European Qualifications Framework implementation*. Retrieved from European Parliament website: http://www.europarl.europa.eu/RegData/etudes/etudes/join/2012/474537/IPOL-CULT_ET%282012%29474537_EN.pdf

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2), 101-133.

Bruner, J. S. (1961). The act of discovery. *Harvard educational review, 31*, 21-32.

Bucchi, M. (2008). Of deficits, deviations and dialogues: Theories of public communication of science. In M. Bucchi, & B. Trench (Eds.), *Handbook of public communication of science and technology* (pp. 71-90). London: Routledge.

Bucchi, M., & Saracino, B. (2016). "Visual Science Literacy". *Science Communication, 38*, 812-819. doi:10.1177/1075547016677833

Buddle, E. A., Bray, H. J., & Ankeny, R. A. (2018). Why would we believe them? Meat consumers' reactions to online farm animal welfare activism in Australia. *Communication Research and Practice, 4*(3), 246-260. doi:10.1080/22041451.2018.1451209

Bulunuz, M., & Jarrett, O. S. (2010). Developing an Interest in Science: Background Experiences of Preservice Elementary Teachers. *International Journal of Environmental and Science Education*, 5(1), 65-84.

Burns, M., & Medvecky, F. (2016). The disengaged in science communication: How not to count audiences and publics. *Public Understanding of Science*, *27*(2), 118-130.

Burns, T. W., O'Connor, D. J., & Stocklmayer, S. M. (2003). Science communication: a contemporary definition. *Public Understanding of Science, 12*(2), 183-202.

Buskirk, T. D., Saunders, T., & Michaud, J. (2015). Are sliders too slick for surveys? An experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys. *methods, data, analyses, 9*(2), 229-260. doi:10.12758/mda.2015.013

Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine, 26*(19), 3661-3675.

Cape, P. (2009). *Slider scales in online surveys.* Paper presented at the CASRO Survey Panel Conference, New Orleans, LA.

Cardiel, C., & Pattison, S. (2014). *Science on the Move: Front-end evaluation report*. Retrieved from https://www.omsi.edu/evaluationreports

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education, 42*(12), 1150-1152.

Carver, C. S. (2001). Affect and the functional bases of behavior: On the dimensional structure of affective experience. *Personality and Social Psychology Review, 5*(4), 345-356.

Cervetti, G., & Hiebert, E. H. (2015). Knowledge, literacy, and the Common Core. *Language Arts, 92*(4), 256-269.

Ceuppens, S., Deprez, J., Dehaene, W., & De Cock, M. (2018). Design and validation of a test for representational fluency of 9th grade students in physics and mathematics: The case of linear functions. *Physical Review Physics Education Research, 14*(2), 1-19.

Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain, 78*(1), 27-37.

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. New York, NY: Springer-Verlag.

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). *Data cleaning: Overview and emerging challenges.* Paper presented at the International Conference on Management of Data, San Francisco, CA.

Cigrik, E., & Ozkan, M. (2015). The investigation of the effect of visiting science center on scientific process skills. *Procedia-Social and Behavioral Sciences, 197*, 1312-1316.

Clark, E., & Araki, K. (2011). Text normalization in social media: progress, problems and applications for a pre-processing system of casual English. *Procedia-Social and Behavioral Sciences, 27*, 2-11.

Clason, D. L., & Dormody, T. J. (1994). Analyzing Data Measured By Individual Likert-Type Items. *Journal of Agricultural Education, 35*, 31-35. doi:10.5032/jae.1994.04031

Cleophas, T. J., & Zwinderman, A. H. (2016). Clinical Data Analysis on a Pocket Calculator: Understanding the Scientific Methods of Statistical Reasoning and Hypothesis Testing. Cham, Switzerland: Springer International.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829-836.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596-610.

Cobern, W. W. (1993). Contextual Constructivism: The Impact of Culture on the Learning and Teaching of Science. In K. Tobin (Ed.), *The practice of constructivism in science education* (pp. 67-86). New York, NY: Routledge.

Cofer, D. A., Jr. (2000). *Informal Learning in the Workplace: A Brief Review of Practice and Application*. Retrieved from https://eric.ed.gov/?id=ED441160

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155-159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.) New York, NY: Routledge.

Colliver, Y., & Fleer, M. (2016). ‘I already know what I learned’: young children's perspectives on learning through play. *Early Child Development and Care, 186*(10), 1559-1570.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in cognitive science, 8*(1), 160-179.

Corden, A., & Sainsbury, R. (2006). *Using verbatim quotations in reporting qualitative social research: Researchers' views*. York, UK: University of York.

Cortassa, C. (2016). In science communication, why does the idea of a public deficit always return? The eternal recurrence of the public deficit. *Public Understanding of Science, 25*(4), 447-459. doi:10.1177/0963662516629745

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Coughlin, S. S., Aliaga, P., Barth, S., Eber, S., Maillard, J., Mahan, C., & Williams, M. (2011). The effectiveness of a monetary incentive on response rates in a survey of recent US veterans. *Survey Practice, 4*(1), 1-8.

Couper, M. P. (2007). Issues of representation in eHealth research (with a focus on web surveys). *American Journal of Preventive Medicine, 32*(5), S83-S89.

Crawford, K., Briggs, K. K., Rodkey, W. G., & Steadman, J. R. (2007). Reliability, validity, and responsiveness of the IKDC score for meniscus injuries of the knee. *Arthroscopy: The Journal of Arthroscopic & Related Surgery, 23*(8), 839-844.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and Conducting Mixed Methods Research*. Los Angeles, CA: SAGE Publications.

Crone, E. A., & Ridderinkhof, K. R. (2011). The developing brain: from theory to neuroimaging and back. *Developmental Cognitive Neuroscience, 1*(2), 101-109.

Crystal, D. (2008). *Txtng : The gr8 db8* [eBook]. Retrieved from https://ebookcentral-proquest-com.ezproxy.otago.ac.nz

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper Perennial.

Csikszentmihalyi, M., & Hermanson, K. (1995). Intrinsic motivation in museums: Why does one want to learn? In J. H. Falk & L. Dierking (Eds.), *Public institutions for personal learning: Establishing a research agenda* (pp. 67-77). Washington, DC: American Association of Museums.

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1997). *Talented Teenagers: The Roots of Success and Failure*. Cambridge, UK: Cambridge University Press.

Cumming, G. (2013). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* [eBook] doi:10.4324/9780203807002

Danesi, M. (2016). *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. London, UK: Bloomsbury Publishing.

Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: a Foundation for Analysis in the Health Sciences* (11th ed.). Hoboken, NJ: Wiley.

Davies, S. R., Halpern, M., Horst, M., Kirby, D., & Lewenstein, B. (2019). Science stories as culture: experience, identity, narrative and emotion in public communication of science. *Journal of Science Communication, 18*(5), A01.

Davis, S. K., Thompson, J. L., & Schweizer, S. E. (2012). Innovations in On-Site Survey Administration: Using an iPad Interface at National Wildlife Refuges and National Parks. *Human Dimensions of Wildlife, 17,* 282-294. doi:10.1080/10871209.2012.673242

Dawson, E. (2012). *Non-participation in public engagement with science: A study of four socio-economically disadvantaged, minority ethnic groups* (Doctoral dissertation). King's College London (University of London). Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.628127

De Cruz, H. (2006). Why are some numerical concepts more successful than others? An evolutionary perspective on the history of number concepts. *Evolution and Human Behavior, 27*(4), 306-323.

De Gruyter Saur. (2017). *Museums of the world 2017* (24th ed.). Munich, Germany: De Gruyter Saur.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*(2), 202-209.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review, 20*(4), 631-642.

Dean, K., Walker, Z., & Jenkinson, C. (2018). Data quality, floor and ceiling effects, and test-retest reliability of the Mild Cognitive Impairment Questionnaire. *Patient Related Outcome Measures, 9*, 43-47.

Dearing, E., & Hamilton, L. C. (2006). Contemporary advances and classic advice for analyzing mediating and moderating variables. *Monographs of the society for research in child development, 71*(285), 88-104. doi:10.1111/j.1540-5834.2006.00406.x

DeBacker, T. K., & Nelson, R. M. (2000). Motivation to learn science: Differences related to gender, class type, and ability. *The Journal of Educational Research, 93*(4), 245-254.
DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37(6), 582-601.
Derrick, B., & White, P. (2017). Comparing two samples from an individual Likert question. *International Journal of Mathematics and Statistics, 18*(3).
DeWitt, J., & Osborne, J. (2010). Recollections of Exhibits: Stimulated-recall interviews with primary school children about science centre visits. *International Journal of Science Education, 32*(10), 1365-1388.
Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science, 40*(3), 434-449.
Diamond, J. (1986). The behavior of family groups in science museums. *Curator: The museum journal, 29*(2), 139-154.
Diamond, J., Horn, M., & Uttal, D. H. (2016). *Practical evaluation guide: Tools for museums and other informal educational settings*. Lanham, MD: Rowman & Littlefield.
Dicks, B. (2013). Interacting with … what? Exploring children's social and sensory practices in a science discovery centre. *Ethnography and Education, 8*, 301-322. doi:10.1080/17457823.2013.792677
Diggle, P. J., & Chetwynd, A. G. (2011). *Statistics and Scientific Method: an Introduction for Students and Researchers*. Oxford, UK: Oxford University Press.
Diseth, Å., Meland, E., & Breidablik, H. J. (2014). Self-beliefs among students: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences, 35*, 1-8.
Dou, D., Wang, H., & Liu, H. (2015). *Semantic data mining: A survey of ontology-based approaches.* Paper presented at the 9th International Conference on Semantic Computing, Anaheim, CA.
Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences, 114*(36), 9587-9592. doi: 10.1073/pnas.1704882114
Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Vol. 44, pp. 247-296). San Diego, CA: Elsevier.
Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest, 5*(3), 69-106.
Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology, 34*(9), 917-928.
Eaton, S. E. (2010). *Formal, non-formal and informal learning: The case of literacy, essential skills, and language learning in Canada*. Retrieved from http://www.en.copian.ca/library/research/eaton/eaton.pdf
Eccles, J., Templeton, J., Barber, B., & Stone, M. (2003). Adolescence and emerging adulthood: The critical passage ways to adulthood. In M. Bornstein, & L. Davidson (Eds.), *Well-being: Positive development across the life course* (pp. 383-406). Mahwah, NJ: Lawrence Erlbaum Associates.

Ecsite-uk. (2008). *Inspiration, engagement and learning: The value of Science & Discovery Centres in the UK, working towards a benchmarking framework*. Bristol, England: Ecsite-uk.

Edwards, P. J., Roberts, I., Clarke, M. J., DiGuiseppi, C., Wentz, R., Kwan, I., … Pratap, S. (2009). Methods to increase response to postal and electronic questionnaires. *Cochrane database of systematic reviews*, *8*(3), MR000008.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Eghbal-Azar, K., Merkt, M., Bahnmueller, J., & Schwan, S. (2016). Use of digital guides in museum galleries: Determinants of information selection. *Computers in Human Behavior, 57*, 133-142.

Ellenbogen, K. M., Luke, J. J., & Dierking, L. D. (2004). Family learning research in museums: An emerging disciplinary matrix? *Science Education*, *88*(S1), S48-S58.

Emereole, H. U. (2009). Learners' and teachers' conceptial knowledge of science processes: The case of Botswana. *International Journal of Science and Mathematics Education, 7*(5), 1033-1056.

Enideo. (2019). WordItOut [program]. Retrieved from https://worditout.com.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications, 14*(1), 153-158.

Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology*, *16*(2), 171-190.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, *5*(6), 00149.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1-4.

European Parliment Council. (2008). *Recommendation of the European Parliment and of the the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning* (2008/C 111/01). Retrieved from https://ec.europa.eu/ploteus/sites/eac-eqf/files/journal_en.pdf

Fagerland, M. W., Lydersen, S., & Laake, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical methods in medical research, 24*(2), 224-254.

Falk, J. H. (2005). Free-choice environmental learning: framing the discussion. *Environmental education research*, *11*(3), 265-280.

Falk, J. H., Dierking, L., & Foutz, S. (2007). *In principle, in practice: Museums as learning institutions*. Lanham, MD: Rowman & Littlefield.

Falk, J. H., & Dierking, L. D. (2004). The contextual model of learning (2000). In G. Anderson (Ed.), *Reinventing the Museum: Historical and Contemporary Perspectives on the Paradigm Shift* (pp. 139-142). Lanham, MD: AltaMira Press.

Falk, J. H., & Dierking, L. D. (2012). *The museum experience revisited*. New York, NY: Routledge.

Falk, J. H., & Dierking, L. D. (2016). *The museum experience revisited*. New York, NY: Routledge.

Falk, J. H., Dierking, L. D., Swanger, L. P., Staus, N., Back, M., Barriault, C., Catalao, C., Chambers, C., Chew, L., Dahl, S. A., Falla, S., Gorecki, B., Lau, T., Lloyd, A., Martin, J., Santer, J., Singer, S., Solli, A., Trepanier, G., Tyystjärvi, P. (2016). Correlating science center use with adult science literacy: an international, cross-institutional study. *Science Education: Science Learning in Everyday Life, 100 (5),* pp. 849-876.

Falk, J. H., & Dierking, L. D. (2018). *Learning from Museums* (2nd ed.). Lanham, MD: Rowman & Littlefield.

Falk, J. H., & Needham, M. D. (2011). Measuring the impact of a science center on its community. *Journal of Research in Science Teaching, 48*(1), 1-12. doi:10.1002/tea.20394

Falk, J. H., & Needham, M. D. (2013). Factors contributing to adult knowledge of science and technology. *Journal of Research in Science Teaching, 50*(4), 431-452. doi:10.1002/tea.21080

Falk, J. H., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education, 89*(5), 744-778. doi:10.1002/sce.20078

Falk, J. H., Storksdieck, M., & Dierking, L. D. (2007). Investigating public science interest and understanding: evidence for the importance of free-choice learning. *Public Understanding of Science, 16*(4), 455-469. doi:10.1177/0963662506064240

Fan, X., Miller, B. C., Park, K.-E., Winward, B. W., Christensen, M., Grotevant, H. D., & Tai, R. H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods, 18*(3), 223-244.

Farrell, B., & Medvedeva, M. (2010). *Demographic transformation and the future of museums*. Washington, DC: The AAM Press.

Fayn, K., Silvia, P. J., Erbas, Y., Tiliopoulos, N., & Kuppens, P. (2018). Nuanced aesthetic emotions: Emotion differentiation is related to knowledge of the arts and curiosity. *Cognition and Emotion, 32*(3), 593-599.

Feher, E. (1990). Interactive museum exhibits as tools for learning: explorations with light. *International Journal of Science Education, 12*(1), 35-49. doi:10.1080/0950069900120104

Feher, E., & Meyer, K. R. (1992). Children's conceptions of color. *Journal of Research in Science Teaching, 29*(5), 505-520.

Feher, E., & Rice, K. (1988). Shadows and anti-images: Children's conceptions of light and vision. II. *Science Education, 72*(5), 637-649.

Feinstein, N. (2011). Salvaging science literacy. *Science Education, 95*(1), 168-185. doi:10.1002/sce.20414

Fencl, H. S. (2010). Development of students' critical-reasoning skills through content-focused activities in a general education course. *Journal of College Science Teaching, 39*(5), 56-62.

Fender, J. G., & Crowley, K. (2007). How parent explanation changes what children learn from everyday scientific thinking. *Journal of Applied Developmental Psychology, 28*(3), 189-210.

Fenichel, M., & Schweingruber, H. (2010). *Surrounded by Science: Learning science in informal environments*. Washington, DC: The National Academies Press.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research, 40*(5), 532-538.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Los Angeles, CA: SAGE Publications.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. Cambridge, UK: Cambridge University Press.

Fors, V. (2006). *The missing link in learning in science centres* (Doctoral dissertation). Luleå. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-25684

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion, 13*(2), 280.

Fowler, F. J., Jr. (2013). *Survey research methods* (5th ed.). Los Angeles, CA: SAGE Publications.

Fraser, B. J. (1978). Development of a test of science-related attitudes. *Science Education, 62*(4), 509-515.

Fraser, B. J. (1981). *TOSRA: Test of science-related attitudes.* Melbourne, Australia: Australian Council for Educational Research.

Frechtling Westat, J., Mark, M. M., Rog, D. J., Thomas, V., Frierson, H., Hood, S., & Hughes, G. (2010). *The 2010 User-Friendly Handbook for Project Evaluation*. Retrieved from https://www.informalscience.org/sites/default/files/TheUserFriendlyGuide.pdf

Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, *5*(1), 20-33.

Frehill, L. M., & Pelczar, M. (2018). *Data File Documentation: Museum Data Files, FY 2018 Release*. Washington, DC. Retrieved from https://www.imls.gov/sites/default/files/museum_data_file_documentation_and_users_guide.pdf

Frey, B. B. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Los Angeles, CA: SAGE Publications.

Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*(5), 873-884.

Friedman, A. J. (Ed.). (2008). *Framework for Evaluating Impacts of Informal Science Education Projects*. Retrieved from https://www.informalscience.org/framework-evaluating-impacts-informal-science-education-projects

Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management, 9*(3), 114-123.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18.

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft, 69*(2), 195-210.

Füchslin, T., Schäfer, M. S., & Metag, J. (2018). A short survey instrument to segment populations according to their attitudes toward science. Scale development, optimization and assessment. *Environmental Communication, 12*(8), 1095-1108.

Ganeb, M. D., & Morales, M. P. E. (2018). Science fluency in primary school: Student transition from Filipino to English language learning. *Issues in Educational Research, 28*(3), 596-612.

Ganu, G., Elhadad, N., & Marian, A. (2009). *Beyond the stars: improving rating predictions using review text content.* Paper presented at the 12th International Workshop on the Web and Databases, Providence, RI.

García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining*. Cham, Switzerland: Springer.

Gauchat, G. (2011). The cultural authority of science: Public trust and acceptance of organized science. *Public Understanding of Science, 20*(6), 751-770. doi:10.1177/0963662510365246

Gazi, A. (2014). Exhibition ethics-An overview of major issues. *Journal of Conservation and Museum Studies, 12*(1), 1-10.

Gee, J. P. (2012). *Situated language and learning: A critique of traditional schooling*. New York, NY: Routledge.

George, D., & Mallery, P. (2009). *SPSS for Windows Step by Step: a Simple Guide and Reference, 17.0 update* (10th ed.). Boston. MA, USA: Allyn & Bacon.

Gibbs, G. R. (2018). *Analyzing Qualitative Data* (2nd ed.). London, UK: SAGE Publications.

Giduthuri, J. G., Maire, N., Joseph, S., Kudale, A., Schaetti, C., Sundaram, N., … Weiss, M. G. (2014). Developing and validating a tablet version of an illness explanatory model interview for a public health survey in pune, india. *PLoS One, 9*(9), e107374. doi:10.1371/journal.pone.0107374

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*(5), 587-606.

Gilman, B. I. (1916). Museum fatigue. *The Scientific Monthly, 2*(1), 62-74.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, OH.

Gluckman, P. (2011). *Looking ahead: Science education for the twenty-first century*. Retrieved from Office of the Prime Minister's Chief Science Advisor website https://www.pmcsa.org.nz/wp-content/uploads/Looking-ahead-Science-education-for-the-twenty-first-century.pdf

Gondwe, M., & Longnecker, N. (2015). Objects as Stimuli for Exploring Young People's Views about Cultural and Scientific Knowledgescientific knowledge. *Science, Technology, & Human Values, 40*, 766-792. doi:10.1177/0162243915577452

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, *11*(4), 364-377.

Grack Nelson, A., & Cohn, S. (2015). Data collection methods for evaluating museum programs and exhibitions. *Journal of Museum Education, 40*(1), 27-36.

Griffin, J., Kelly, L., Savage, G., & Hatherly, J. (2005). Museums actively researching visitors experiences and learning (MARVEL): a methodological study. *Open Museum Journal*, *7*(1), 1-19.

Griffin, J. M., Simon, A. B., Hulbert, E., Stevenson, J., Grill, J. P., Noorbaloochi, S., & Partin, M. R. (2011). A comparison of small monetary incentives to convert survey non-respondents: a randomized control trial. *BMC Medical Research Methodology*, *11*(1), 81.

Grigorovitch, A. (2014). Children's misconceptions and conceptual change in Physics Education: the concept of light. *Journal of Advances in Natural Sciences*, *1*(1), 34-39.

Grinell, S. (1988). Science Centers Come of Age. *Issues in Science and Technology, 4*(3), 70-75.

Grissom, R. J., & Kim, J. J. (2005). *Effect Sizes for Research: A Broad Practical Approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Groff, A., Lockhart, D., Ogden, J., & Dierking, L. D. (2005). An exploratory investigation of the effect of working in an environmentally themed facility on the conservation-related knowledge, attitudes and behavior of staff. *Environmental Education Research*, *11*(3), 371-387.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). Validity and Reliability (Credibility and Dependability) in Qualitative Research and Data Analysis. In G. Guest, K. M. MacQueen, & E. E. Namey, *Applied Thematic Analysis* (pp. 79-106). London: Sage Publications.

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: relations between parents' input and children's cardinal-number knowledge. *Developmental Science*, *14*(5), 1021-1032.

Guo, Y., Kopec, J. A., Cibere, J., Li, L. C., & Goldsmith, C. H. (2016). Population survey features and response rates: a randomized experiment. *American Journal of Public Health*, *106*(8), 1422-1426.

Gürel, E., & Tat, M. (2017). SWOT analysis: a theoretical review. *Journal of International Social Research*, *10*(51).

Gutwill, J. P. (2018). Science Self-Efficacy and Lifelong Learning: Emerging Adults in Science Museums. *Visitor Studies, 21*(1), 31-56.

Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). *School Science and Mathematics, 114*(6), 271-279.

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health, 11*, 322-333. doi:10.1111/j.1524-4733.2007.00231.x

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis: Pearson* (new int. ed.). Harlow, UK: Pearson Education.

Halim, L., Rahman, N. A., Zamri, R., & Mohtar, L. (2018). The roles of parents in cultivating children's interest towards science learning and careers. *Kasetsart Journal of Social Sciences, 39*(2), 190-196.

Hall, L., Hume, C., & Tazzyman, S. (2016). *Five degrees of happiness: Effective smiley face likert scales for evaluating with children.* Paper presented at the Proceedings of the The 15th International Conference on Interaction Design and Children, New York, NY.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research, 7*(1), 1-20.

Ham, S. (2016). *Interpretation: Making a difference on purpose*. Golden, CO: Fulcrum Publishing.

Hampp, C., & Schwan, S. (2015). The role of authentic objects in museums of the history of science and technology: Findings from a visitor study. *International Journal of Science Education, Part B, 5*(2), 161-181.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.).: Waltham, MA: Morgan Kaufmann.

Han, P., Shen, S., Wang, D., & Liu, Y. (2012). *The influence of word normalization in English document clustering.* Paper presented at the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China.

Harré, N. (2011). *Psychology for a better world: Strategies to inspire sustainability.* Auckland, NZ: University of Auckland.

Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research, 39*(6), 701-723. doi: 10.1177/0093650211416646

Hassan, H., & Menezes, A. (2013). *Social text normalization using contextual graph random walks.* Paper presented at the Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria.

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon, UK: Routledge.

Hauan, N. P., DeWitt, J., & Kolstø, S. D. (2017). Proposing an evaluation framework for interventions: focusing on students' behaviours in interactive science exhibitions. *International Journal of Science Education, Part B, 7*(2), 103-120.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (2nd ed.). New York, NY: The Guilford Press.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures, 1*(1), 77-89.

Hayes, J. R., & Hatch, J. A. (1999). Issues in measuring reliability: Correlation versus percentage of agreement. *Written Communication, 16*(3), 354-367.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based Nursing, 18*(3), 66-67.

Health Promotion Agency. (2019). *SunSmart*. Retrieved from https://www.sunsmart.org.nz/

Heath, C., & Vom Lehn, D. (2008). Configuring 'Interactivity': Enhancing Engagement in Science Centres and Museums. *Social Studies of Science, 38*(1), 63-91.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press.

Heimlich, J. E., & Horr, E. E. T. (2010). Adult learning in free-choice, environmental settings: What makes it different? *New Directions for Adult and Continuing Education, 2010*(127), 57-66.

Hennink, M. M. (2013). *Focus Group Discussions*. New York, NY: Oxford University Press.

Herd, J. (2005). Education. In *Cracks in a glass ceiling: New Zealand women 1975-2004* (pp. 74-88). Dunedin, New Zealand: NZ Federation of Graduate Women (Otago Branch).

Hernández, S., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación* [Investigation Methodology] (6th ed.). Mexico City, Mexico: McGraw-Hill.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.

Hikuroa, D. (2017). Mātauranga Māori—the ūkaipō of knowledge in New Zealand. *Journal of the Royal Society of New Zealand, 47*(1), 5-10.

Hill, M., & Sharma, M. D. (2015). Students' representational fluency at university: A cross-sectional measure of how multiple representations are used by physics students using the representational fluency survey. *Eurasia Journal of Mathematics, Science and Technology Education, 11*(6), 1633-1655.

Hill, M., Sharma, M. D., & Johnston, H. (2015). How online learning modules can improve the representational fluency and conceptual understanding of university physics students. *European Journal of Physics, 36*(4), 045019.

Hill, M., Sharma, M. D., O'Byrne, J., & Airey, J. (2014). Developing and evaluating a survey for representational fluency in science. *International Journal of Innovation in Science and Mathematics Education, 22*(6), 22-42.

Hirschman, E. C. (1986). The effect of verbal and pictorial advertising stimuli on aesthetic, utilitarian and familiarity perceptions. *Journal of Advertising, 15*(2), 27-34.

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement, 75*(3), 365-388.

Hodzi, R. (1992). Relationship of attitudes toward self, family, and school with attitude toward science among secondary school students in Zimbabwe. *Zimbabwe Journal of Educational Research, 4*(3), 294-306.

Hohenstein, J., & Tran, L. U. (2007). Use of questions in exhibit labels to generate explanatory conversation among science museum visitors. *International Journal of Science Education, 29*(12), 1557-1580.

Hohwü, L., Lyshol, H., Gissler, M., Jonsson, S. H., Petzold, M., & Obel, C. (2013). Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. *Journal of Medical Internet Research, 15*, e173. doi:10.2196/jmir.2595

Holbrook, J., & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental and Science Education, 4*(3), 275-288.

Hollos, M. (1975). Logical operations and role-taking abilities in two cultures: Norway and Hungary. *Child Development*, 46, 638-649.

Hollos, M., & Cowan, P. A. (1973). Social isolation and cognitive development: Logical operations and role-taking abilities in three Norwegian social settings. *Child Development*, *44*(3), 630-641.

Holmes, J. A. (2011). Informal learning: Student achievement and motivation in science through museum-based learning. *Learning Environments Research, 14*(3), 263-277.

Holstermann, N., Grube, D., & Bögeholz, S. (2010). Hands-on activities and their influence on students' interest. *Research in Science Education, 40*(5), 743-757.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17,* 63-84.

Honomichl, R. D., & Chen, Z. (2012). The role of guidance in children's discovery learning. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(6), 615-622.

Hooper-Greenhill, E. (2002). *Developing a scheme for finding evidence of the outcomes and impact of learning in museums, archives and libraries: the conceptual framework*. Retrieved from Leicester: https://www2.le.ac.uk/departments/museumstudies/rcmg/projects/lirp-1-2/LIRP%20analysis%20paper%201.pdf

Housen, A. (2007). Art viewing and aesthetic development: Designing for the viewer. In P. Villenueve (Ed.), *From periphery to center: Art museum education in the 21st century* (pp. 172-179). Reston, VA: NAEA.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of school psychology*, *49*(5), 505-528.

Illeris, K. (2018). *Contemporary theories of learning: learning theorists... in their own words*. London, UK: Routledge.

International Association for the Evaluation of Educational Achievement. (1994). *Third International Mathematics and Science Study: Supplement 1*. Retrieved from IEA website: https://timss.bc.edu/timss1995i/database/UG1_Sup1.pdf

International Association for the Evaluation of Educational Achievement. (1998). IEA *Third International Mathematics and Science Study - Repeat: Student questionnaire main survey*. Retrieved from IEA website: https://timssandpirls.bc.edu/timss1999i/pdf/BM2_StudentG.pdf

International Association for the Evaluation of Educational Achievement. (2002). *IEA Trends in International Mathematics and Science Study 2003: Main survey student questionnaire grade 4.* Retrieved from IEA website: https://timss.bc.edu/timss2003i/PDF/T03_Student_4.pdf

International Association for the Evaluation of Educational Achievement. (2018a). *IEA Trends in International Mathematics and Science Study 2003: Main survey student questionnaire grade 4*. Retrieved from IEA website: https://nces.ed.gov/timss/pdf/T19_GR4_StudentQ_USA_Questionnaire.pdf

International Association for the Evaluation of Educational Achievement. (2018b). *Trends in International Mathematics and Science Study 2019: Student questionnaire grade 8*. Retrieved from IEA website: https://nces.ed.gov/timss/pdf/T19_GR8_StudentQ_USA_Questionnaire.pdf

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The effect of interviewer personality, skills and attitudes on respondent co-operation with face-to-face surveys. *Survey Research Methods, 7*, 1-15.

Jacobsen, J. W. (2016). *Measuring Museum Impact and Performance: Theory and Practice*. Lanham, MD: Rowman & Littlefield.

Jacoby, J., & Matell, M. S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research, 8*(4), 495-500.

Jacoby, W. G. (2000). Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies, 19*(4), 577-613.

Jančić, O. C. (2016). Rights of the Child in a Changing World. In O. C. Jančić (Ed.), *The rights of the child in a changing world: 25 Years after the UN Convention of the Rights of the Child* (pp. 1-36). Cham, Switzerland: Springer International.

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology, 41*, 13-24. doi:10.1016/j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11-21.

Jarvis, T., & Pell, A. (2002). Effect of the challenger experience on elementary children's attitudes to science. *Journal of Research in Science Teaching, 39*(10), 979-1000.

Jarvis, T., & Pell, A. (2005). Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Centre. *Journal of Research in Science Teaching, 42*(1), 53-83.

Jensen, E., & Buckley, N. (2014). Why people attend science festivals: Interests, motivations and self-reported benefits of public engagement with research. *Public Understanding of Science, 23*(5), 557-573.

Jivani, A. G. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl, 2*(6), 1930-1938.

Jones, J., & Smith, J. (2017). Ethnography: challenges and opportunities. *Evidence Based Nursing, 20*, 98-100.

Jones, J. L., & Sinclair, B. (2011). Assessment on the go: surveying students with an iPad. *Library, 2*, 22-35.

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education, 84*(2), 180-192.

Joseph, G. G. (2008). A brief history of zero. *Iranian Journal for the History of Science, 6*, 37-48.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change, 2*, 732-735. doi:10.1038/nclimate1547

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*(1), 111-117.

Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education, 68*, 545-556.

Kamolpattana, S., Chen, G., Sonchaeng, P., Wilkinson, C., Willey, N., & Bultitude, K. (2015). Thai visitors’ expectations and experiences of explainer interaction within a science museum context. *Public Understanding of Science, 24*(1), 69-85.

Kano, A., & Read, J. (2012). *Interchangeability of computer and paper based questionnaires in gathering computer experience data from young children.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX.

Kanyangarara, T., Mayberry, B. A., Pai, A., & Shanahan, M. (2012). *Gender analysis in education*. Retrieved from http://www.ungei.org/files/genderanalysisineducation.pdf

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968.

Kay, M., & Röscheisen, M. (1993). Text-translation alignment. *Computational linguistics, 19*(1), 121-142.

Kear, D. J., Coffman, G. A., McKenna, M. C., & Ambrosio, A. L. (2000). Measuring attitude toward writing: A new tool for teachers. *The Reading Teacher*, *54*(1), 10-23.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *International Journal of Public Opinion Quarterly, 70*(5), 759-779.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1-24.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in cognitive science, 2*(2), 285-305.

Kilgarriff, A. (2007). Googleology is bad science. *Computational linguistics, 33*(1), 147-151.

Kim, H.-Y. (2017). Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics, 42*(2), 152-155.

Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis, IN: John Wiley & Sons.

Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education*, *29*(7), 871-893.

King, J. D., Buolamwini, J., Cromwell, E. A., Panfel, A., Teferi, T., Zerihun, M., … Emerson, P. M. (2013). A Novel Electronic Data Collection System for Large-Scale Surveys of Neglected Tropical Diseases. *PLoS One, 8,* e74570. doi:10.1371/journal.pone.0074570

Kitayama, S., & Cohen, D. (Eds.). (2010). *Handbook of Cultural Psychology* (1st ed.). New York, NY: Guilford Press.

Koballa, T. R., Jr., & Glynn, S. M. (2010). Attitudinal and motivational constructs in science learning. In S. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (2010 reprint, pp. 75-102). New York, NY: Routledge.

Kohut, A., Keeter, S., Doherty, C., Dimock, M., & Christian, L. (2012). *Assessing the representativeness of public opinion surveys.* Washington, DC: Pew Research Center.

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25*, 67-72.

Koppman, S., Cain, C. L., & Leahey, E. (2015). The joy of science: disciplinary diversity in emotional accounts. *Science, Technology, & Human Values, 40*(1), 30-70.

Koro-Ljungberg, M. (2015). *Reconceptualizing Qualitative Research: Methodologies Without Methodology*. London, UK: SAGE Publications.

Koster, C. H. A. (2004). *Transducing text to multiword units.* Paper presented at the 4th International Conference on Language Resources and Evaluation. Lisbon, Portugal.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science, 1*(2), 111-117.

Kottasz, R. (2006). Understanding the influences of atmospheric cues on the emotional responses and behaviours of museum visitors. *Journal of Nonprofit & Public Sector Marketing, 16*(1-2), 95-121.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science, 328*(5977), 456-459.

Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education, 33*(1), 27-50. doi:10.1080/09500693.2010.518645

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. Retrieved from http://repository.upenn.edu/asc_papers/43

Krishnamurthi, A., & Rennie, L. J. (2012). *Informal science learning and education: definition and goals*. Retrieved from http://repository.upenn.edu/asc_papers/43

Kromann, C., Bohnstedt, C., Jensen, M., & Ringsted, C. (2010). The testing effect on skills learning might last 6 months. *Advances in Health Sciences Education, 15*(3), 395-401.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Michael Hanemann, W., Kopp, R. J., . . . Kerry Smith, V. (2002). The impact of" no opinion" response options on data quality: non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66*(3), 371-403.

Krueger, R. A., & Casey, M. A. (2000). *Focus Groups: A Practical Guide for Applied Research* (3rd ed.). Los Angeles, CA: SAGE Publications.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151-160.

Kurtz-Costes, B., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender stereotypes about mathematics and science and self-perceptions of ability in late childhood and early adolescence. *Merrill-Palmer Quarterly,* 54(3), 386-409.

La Belle, T. J. (1982). Formal, nonformal and informal education: A holistic perspective on lifelong learning. *International Review of Education, 28*(2), 159-175.

Labudde, P., Herzog, W., Neuenschwander, M. P., Violi, E., & Gerber, C. (2000). Girls and physics: Teaching and learning strategies tested by classroom interventions in grade 11. *International Journal of Science Education, 22*(2), 143-157.

Ladwig, P., Dalrymple, K. E., Brossard, D., Scheufele, D. A., & Corley, E. A. (2012). Perceived familiarity or factual knowledge? Comparing operationalizations of scientific understanding. *Science and Public Policy, 39*(6), 761-774.

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education, 52*(5), 537-553.

Lantz, B. (2013). Equidistance of Likert-type scales and validation of inferential methods using experiments and simulations. *The Electronic Journal of Business Research Methods, 11*(1), 16-28.

Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *The Lancet, 378*(9790), 526-535.

Latchem, C. (2014). Informal learning and non-formal education for development. *Journal of Learning for Development, 1*(1). Retrieved from https://jl4d.org/index.php/ejl4d/article/view/6/6

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, UK: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lavrakas, P. J. (Ed.) (2008). *Encyclopedia of survey research methods.* Thousand Oaks, CA: Sage Publications.

Laz, C. (1998). Act your age. *Sociological Forum, 13*(1), 85-113.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., …Rothschild, D. (2018). The science of fake news. *Science, 359*(6380), 1094-1096. doi:10.1126/science.aao2998

Le Clus, M. A. (2011). Informal learning in the workplace: A review of the literature. *Australian Journal of Adult Learning*, *51*(2), 355-373.
Lee, J. D. (1998). Which kids can "become" scientists? Effects of gender, self-concepts, and perceptions of scientists. *Social psychology quarterly, 61*(3), 199-219.
Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science achievement. *Science Education*, *97*(4), 574-593.
Leigh, D. (2009). SWOT analysis. In Pershing, J. (Ed.), *Handbook of improving performance in the workplace* (pp. 1089-1108). Pfeiffer.
Leisher, C. (2014). A Comparison of Tablet-Based and Paper-Based Survey Data Collection in Conservation Projects. *Social Sciences, 3*, 264-271. doi:10.3390/socsci3020264
Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. doi: 10.13140/RG.2.2.17823.92329
Lesko, N. (2012). *Act your age!: A Cultural Construction of Adolescence* (2nd ed.). New York, NY: Routledge.
Liberty Science Center. (2019). 12 *Days of Science photo contest*. Retrieved from https://lsc.org/news-and-social/news/want-to-win-an-lsc-membership-enter-our-12-days-of-science-photo-contest
Likert, R. (1932). A technique for the measurement of attitudes. *Archives de Psychologie* [Archives of psychology], 140, 1-55.
Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin, 139*(1), 1-34.
Lin, H.-S., Hong, Z.-R., & Huang, T.-C. (2012). The Role of Emotional Factors in Building Public Scientific Literacy and Engagement with Science. *International Journal of Science Education, 34*(1), 25-42. doi:10.1080/09500693.2010.551430
Linder, C., Östman, L., Roberts, D. A., Wickman, P.-O., Ericksen, G., & MacKinnon, A. (Eds.). (2010). *Exploring the Landscape of Scientific Literacy*. New York, NY: Routledge.
Lindgren-Streicher, A., & Reich, C. A. (2007). Visitor usage of digital and physical artifacts in two museum programs. *Visitor Studies, 10*(2), 152-167.
Lindon, J. (1996). *Growing up: from eight years to young adulthood*. London, UK: National Children’s Bureua Enterprises.
Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.
Liu, C.-C., & Falk, J. H. (2014). Serious fun: Viewing hobbyist activities through a learning lens. *International Journal of Science Education, Part B, 4*(4), 343-355.
Liu, F., Weng, F., & Jiang, X. (2012). *A broad-coverage normalization system for social media language.* Paper presented at the Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju Island, Korea.
Lockwood, A. (2010). Seeding doubt: How sceptics have used new media to delay action on climate change. *Geopolitics, History, and International Relations, 2*(2), 136-164.
Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. New York, NY: Cambridge University Press.
Logi Kristjánsson, Á., Dóra Sigfúsdóttir, I., & Allegrante, J. P. (2010). Health behavior and academic achievement among adolescents: the relative contribution of dietary habits, physical activity, body mass index, and self-esteem. *Health Education & Behavior, 37*(1), 51-64.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research, 28*(4), 587-604.

Longnecker, N. (2016). An integrated model of science communication. *Journal of Science Communication, 15*, 1-12.

Longnecker, N., & Davis, J. (2013). Building the future: the sciences and engineering. In J. Gregory (Ed.), *Seeking Wisdom: A centenary history of the University of Western Australia* (pp. 321-351). Crawley, Australia: UWA Publishing.

Longnecker, N., Elliot, J., & Gondwe, M. (2014). *Inspiring Australia: An evaluation tool for science engagement activities*. Report prepared for the Department of Innovation, Industry, Science, Research and Tertiary Education. Canberra, Australia.

Longnecker, N., & Scott, C. (2018). Challenges of cross-cultural communication in production of a collaborative exhibition: Wai ora, Mauri ora. *Journal of Science Communication, 17*(4), C05.

Lucas, R., Norval, M., Neale, R., Young, A., De Gruijl, F., Takizawa, Y., & Van der Leun, J. (2015). The consequences for human health of stratospheric ozone depletion in association with other environmental factors. *Photochemical & Photobiological Sciences, 14*(1), 53-87.

Lydersen, S., Fagerland, M. W., & Laake, P. (2009). Recommended tests for association in 2× 2 tables. *Statistics in Medicine, 28*(7), 1159-1175.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*(5), 481-498.

MacDonald, E. (2018). Increasing visitor engagement in a contemporary art exhibit through a participatory activity. *Tuhinga,* (29), 165-173.

MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: causes, mechanisms, and procedural remedies. *Journal of Retailing, 88*(4), 542-555.

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy, 10*(2), 198-213.

Malcolm, J., Hodkinson, P., & Colley, H. (2003). The interrelationships between informal and formal learning. *Journal of Workplace Learning, 15*(7/8), 313-318.

Maletic, J. I., & Marcus, A. (2000). *Data Cleansing: Beyond Integrity Analysis.* Paper presented at the Conference on Information Quality (IQ2000), Cambridge, MA.

Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685. doi:10.1080/09500690902792385

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York, NY: Cambridge University Press.

Maple, S. A., & Stage, F. K. (1991). Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal, 28*(1), 37-60.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*(1), 59-77.

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. *APA educational psychology handbook, 1*, 427-458.

Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. *Radiography, 16*(4), e1-e7.

Martin, A. J., Durksen, T. L., Williamson, D., Kiss, J., & Ginns, P. (2016). The role of a museum-based science education program in promoting content knowledge and science motivation. *Journal of Research in Science Teaching, 53*(9), 1364-1384.

Masuda, S., Sakagami, T., Kawabata, H., Kijima, N., & Hoshino, T. (2017). Respondents with low motivation tend to choose middle category: survey questions on happiness in Japan. *Behaviormetrika, 44*(2), 593-605.

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation, 8*(1), 3-30.

Mattar, J. (2018). Constructivism and connectivism in education technology: Active, situated, authentic, experiential, and anchored learning. *Revista Iberoamericana de Educación a Distancia, 21*(2), 201-217.

Maxwell, L. E., & Evans, G. W. (2002). Museums as learning settings: The importance of the physical environment. *Journal of Museum Education, 27*(1), 3-7.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American psychologist, 59*(1), 14-19.

Mayo, E. (2004). *The human problems of an industrial civilization*. New York, NY: Routledge.

Mbewe, S., Chabalengula, V. M., & Mumba, F. (2010). Pre-service teachers' familiarity, interest and conceptual understanding of science process skills. *Problems of Education in the 21st Century, 22*(22), 76-86.

McCallie, E., Bell, L., Lohwater, T., Falk, J. H., Lehr, J. L., Lewenstein, B. V., … Wiehe, B. (2009). *Many Experts, Many Audiences: Public Engagement with Science and Informal Science Education: A CAISE Inquiry Group Report*. Retrieved from http://caise.insci.org/uploads/docs/public_engagement_with_science.pdf

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*(3), 267-277.

McCombs, B. L. (2001). What Do We Know About Learners and Learning? The Learner-Centered Framework: Bringing the Educational System into Balance. *Educational Horizons*, *79*(4), 182-193.

McCubbins, S. (2016). Using a visitor based framework to observe engagement in a children's museum makerspace. (Doctoral dissertation). Illinois State University. Retrieved from https://ir.library.illinoisstate.edu/etd/526

McGahee, T. W., & Ball, J. (2009). How to read and really use an item analysis. *Nurse Educator, 34*(4), 166-171.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica, 23*(2), 143-149.

McKenna-Cress, P., & Kamien, J. (2013). *Creating exhibitions: Collaboration in the planning, development, and design of innovative experiences*. Hoboken, NJ: John Wiley & Sons.

McLeod, A., Pippin, S., & Wong, J. A. (2011). Revisiting the Likert scale: can the fast form approach improve survey research? *International Journal of Behavioural Accounting and Finance, 2*(3-4), 310-327.

McRae, K., Gross, J. J., Weber, J., Robertson, E. R., Sokol-Hessner, P., Ray, R. D., … Ochsner, K. N. (2012). The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social Cognitive and Affective Neuroscience, 7*(1), 11-22.

MedCalc Software (2019). *MedCalc: easy-to-use friendly statistical software* [Software]. Retrieved from www.medcalc.org

Meier, M. D., & Diefenbach, H. (2018). The OECD between political and scientific agendas-a critique of the 2015 PISA gender report. *Gender and Education*. doi:10.1080/09540253.2018.1471198

Mellor, D., & Moore, K. A. (2013). The use of Likert scales with children. *Journal of Pediatric Psychology, 39*(3), 369-379.

Melton, A. W. (1935). *Problems of installation in museums of art* (New Series No. 14). Washington, DC: American Association of Museums.

Moreiera, J., de Carvalho, A., & Horváth, T. (2019). *A General Introduction to Data Analytics.* Hoboken, NJ: Wiley.

Metcalf, J., Alford, K., & Shore, J. (2013). *National audit of Australian science engagement activities 2012.* Retrieved from http://www.econnect.com.au/wp-content/uploads/2013/01/National-Audit-Final-Report_20-01-13-FINAL.pdf

Metzler, J. (2017). Cleaning Up Quotations. *Journal of Appellate Practice & Process, 18*, 143.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin, 105*(1), 156-166.

Miglietta, A. M., Pace, R., & Boero, F. (2011). Evaluating students' comprehension of messages on panels. *Visitor Studies, 14*(2), 209-218.

Mikheev, A. (2000). *Document centered approach to text normalization.* Paper presented at the 23rd ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Miles, R. S., Alt, M., Gosling, D., Lewis, B. N., & Tout, A. F. (Eds.). (1988). *The Design of Educational Exhibits* (2nd ed.). New York, NY: Routledge

Ministry for Culture and Heritage's Cultural Statistics Programme. (2009). *Cultural indicators for New Zealand* Wellington, New Zealand: Ministry for Culture and Heritage. Retrieved from https://mch.govt.nz/sites/default/files/CulturalIndicatorsReport.pdf

Minisitry of Education. (1993). *Science in the New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Ministry of Education. (2014). *Science: Achievement objectives.* Retrieved from http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/Science/Achievement-objectives

Mohtar, L. E., Halim, L., Rahman, N. A., Maat, S. M., Iksan, Z. H., & Osman, K. (2019). A model of interest in stem careers among secondary school students. *Journal of Baltic Science Education, 18*(3), 404-416. doi:10.33225/jbse/19.18.404

Moksony, F. (1990). Small is beautiful. The use and interpretation of R2 in social research. *Szociológiai Szemle, Special Issue*, 130-138.

Mostafa, T. (2019). *Why don't more girls choose to pursue a science career?* Retrieved from https://www.oecd-ilibrary.org/content/paper/02bd2b68-en. doi: 10.1787/02bd2b68-en

Muijtjens, A. M. M., van Mameren, H., Hoogenboom, R. J. I., Evers, J. L. H., & van der Vleuten, C. P. M. (1999). The effect of a 'don't know'option on test scores: Number-right and formula scoring compared. *Medical education, 33*(4), 267-275.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J.F., Preuschoff, C., Erberber, E., … Galia, J. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Retrieved from the TIMSS website: https://timss.bc.edu/timss2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf

Munshi, J. (2014). A method for constructing Likert scales. *SSRN Electronic Journal*, 1-12. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2419366

Murray, J. (2013). Likert data: what to use, parametric or non-parametric? *International Journal of Business and Social Science, 4*(11), 258-264.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology, 142*(2), 71-89.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “×” out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological science, 22*(8), 1058-1066.

Naing, L., Winn, T., & Rusli, B. (2006). Practical issues in calculating the sample size for prevalence studies. *Archives of Orofacial Sciences, 1*, 9-14.

National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, Contexts, and Consequences.* Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures.* Washington, DC: The National Academies Press.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits* (Vol. 32). Washington, DC: The National Academies Press.

National Science Board. (2018). *Science and Engineering Indicators 2018*. Retrieved from https://www.nsf.gov/statistics/2018/nsb20181/

NCSS. (2019a). Chapter 15: Tests for two correlated proportions (McNemar test) [Statistics test]. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/PASS/Tests_for_Two_Correlated_Proportions-McNemar_Test.pdf

NCSS. (2019b). Chapter 151: Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry) [Statstics Test]. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/PASS/Tests_for_Multiple_Correlated_Proportions-McNemar-Bowker_Test_of_Symmetry.pdf

Neuendorf, K. A. (2016). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.

New Zealand Education. (2019). *Education System*. Retrieved from https://www.studyinnewzealand.govt.nz/why-nz/education-system

New Zealand Office of the Prime Minister's Science Advisory Committee. (2011). *Looking ahead: science education for the twenty-first century*. Auckland, New Zealand: Office of the Prime Minister's Science Advisory Committee.

Newcomer, K. E., & Triplett, T. (2015). Using surveys. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 344-382). San Francisco, CA: Jossey-Bass & Pfeiffer Imprints.

Newton, D. P. (2014). *Thinking with Feeling: Fostering productive thought in the classroom.* Abingdon, UK: Routledge.

Nobel, C. (2011). A new model for business: The museum. *Working Knowledge, Harvard Business School, accessed at 04/19/2020*.

Nopparatjamjomras, S., & Chitaree, R. (2009). *Using students' misconceptions of primary coloured lights to design a hands-on coloured light mixer*. Paper presented at the Education and Training in Optics and Photonics Conference, St. Asaph, UK.

Norbury, G., & Jones, C. (2015). Pests controlling pests: does predator control lead to greater European rabbit abundance in A ustralasia? *Mammal Review, 45*(2), 79-87.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625-632. doi:10.1007/s10459-010-9222-y

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

Olitsky, S., & Milne, C. (2012). Understanding engagement in science education: The psychological and the social. In B. Fraser, K. Tobin, & C. J. Campbell (Eds.), *Second International Handbook of Science Education* (pp. 19-33). Dordrecht, Netherlands: Springer.

Olsen, F., Abelsen, B., & Olsen, J. A. (2012). Improving response rate and quality of survey data with a scratch lottery ticket incentive. *BMC Medical Research Methodology, 12*(1), 52.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*(1), 99-114.

Organisation for Economic Cooperation and Development. (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Retrieved from https://www.oecd-ilibrary.org/education/assessing-scientific-reading-and-mathematical-literacy_9789264026407-en

Organisation for Economic Cooperation and Development. (2007). *PISA 2006: Science Competencies for Tomorrow's World, Volume 1: Analysis*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2006_9789264040014-en

Organisation for Economic Cooperation and Development. (2009). *PISA 2006 Technical Report*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2006-technical-report_9789264048096-en

Organisation for Economic Cooperation and Development. (2010). *PISA 2009 Assessment framework: Key Competencies in Reading, Mathematics and Science*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2009-assessment-framework_9789264062658-en

Organisation for Economic Cooperation and Development. (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2012-assessment-and-analytical-framework_9789264190511-en

Organisation for Economic Cooperation and Development. (2013b). *PISA 2012 Results: Ready to Learn (Volume III). Students' Engagement, Drive and Self-Beliefs*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2012-results-ready-to-learn-volume-iii_9789264201170-en

Organisation for Economic Cooperation and Development. (2014). *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science*. Retrieved from https://www.oecd-ilibrary.org/education/pisa-2012-results-what-students-know-and-can-do-volume-i-revised-edition-february-2014_9789264208780-en

Organisation for Economic Cooperation and Development. (2016a). *How does PISA assess science literacy?*. Retrieved from https://www.oecd-ilibrary.org/education/how-does-pisa-assess-science-literacy_5jln4nfnqt7l-en

Organisation for Economic Cooperation and Development. (2016b). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*.

Retrieved from https://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework_9789264255425-en
Organisation for Economic Cooperation and Development. (2018). *PISA 2018 Reading literacy framework*. Retrieved from https://www.iprase.tn.it/documents/20178/344196/Pisa+2018+reading+literacy+framework+_final.pdf
Osborne, J., & Dillon, J. (2008). *Science Education in Europe: Critical Reflections* (Vol. 13). London, UK: The Nuffield Foundation.
Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology, 1*, 220.
Otago Museum. (2015). *Annual Report 2014-2015*. Retrieved from https://otagomuseum.nz/assets/documents/Annual-ReportsPlans-and-Strategic-Plans/OtagoMuseum-1415-AnnualReport.pdf
Otago Museum. (2016). *Annual Report 2015-2016*. Retrieved from https://otagomuseum.nz/assets/documents/Otago-Museum-Annual-Report-2015-2016.pdf
Otago Museum. (2017). *Annual report 2016-2017*. Retrieved from https://otagomuseum.nz/assets/documents/Annual-ReportsPlans-and-Strategic-Plans/OM-Annual-Report-2016-17-EMAIL.pdf
Otago Museum. (2018a). *Annual plan 2018-2019.* Retrieved from https://otagomuseum.nz/assets/documents/OtagoMuseum-2018-19-Annual-Plan.pdf
Otago Museum. (2018b). *Annual report 2017-2018*. Retrieved from https://otagomuseum.nz/assets/Uploads/Otago-Museum-Annual-Report-2017-2018-DIGITAL.pdf
Otago Museum. (2019). *Tūhura*. Retrieved from https://otagomuseum.nz/whats-on/do//tuhura/
Packer, J. (2008). Beyond learning: Exploring visitors' perceptions of the value and benefits of museum experiences. *Curator: The Museum Journal, 51*(1), 33-54.
Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. *Perception, 11*, 239-266.
Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533-544.
Panayides, P. (2013). Coefficient alpha: interpret with caution. *Europe's Journal of Psychology, 9*(4), 687-696.
Parkman, A. (2016). The imposter phenomenon in higher education: Incidence and impact. *Journal of Higher Education Theory and Practice, 16*(1), 51.
Parsons, M. J. (1987). *How we understand art: A cognitive developmental account of aesthetic experience*. New York, NY: Cambridge University Press.
Pasek, J. (2018). It's not my consensus: Motivated reasoning and the sources of scientific illiteracy. *Public Understanding of Science, 27*(7), 787-806. doi:10.1177/0963662517733681
Patel, A. M. (2013). A Research of the Scientific Creativity Components of Higher Secondary School Students. *International Journal of Research in Education, 2*(6), 65-68.
Patiño, M. d. L. (Ed.). (2013). *La divulgación de la ciencia en México desde distintos campos de acción: Visiones, retos y oportunidades* [The dissemination of science in Mexico from different fields of action: Visions, challenges and opportunities]. Mexico City, Mexico: Somedicyt.

Pattison, S. A., Randol, S. M., Benne, M., Rubin, A., Gontan, I., Andanen, E., … Dierking, L. D. (2017). A design-based research study of staff-facilitated family learning at interactive math exhibits. *Visitor Studies, 20*(2), 138-164.
Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York, NY: Guilford Press.
Pear, J. J. (2014). *The science of learning*. New York, NY: Routledge.
Pearce-Morris, J., Choi, S.-W., Roth, V., & Young, R. (2014). Substantive Meanings of Missing Data in Family Research: Does “Don't Know” Matter? *Marriage & Family Review, 50*(8), 665-690.
Peart, B. (1984). Impact of exhibit type on knowledge gain, attitudes, and behavior. *Curator: The museum journal, 27*(3), 220-237.
Pedretti, E. (2002). T. Kuhn meets T. Rex: Critical conversations and new directions in science centres and science museums. *Studies in Science Education,* 37, 1-42.
Pedretti, E., & Navas-Iannini, A. M. (2018). Pregnant pauses: Science museums, schools and a controversial exhibition. In D. Corrigan, C. Buntting, A. Jones, & J. Loughran (Eds.), *Navigating the Changing Landscape of Formal and Informal Science Learning Opportunities* (pp. 31-49). Cham, Switzerland: Springer.
Pekarik, A. J., Schreiber, J. B., & Visscher, N. (2018). Overall Experience Rating-Measuring Visitor Response in Museums. *Curator: The Museum Journal, 61*(2), 353-365.
Peponis, J., Dalton, R. C., Wineman, J., & Dalton, N. (2004). Measuring the effects of layout upon visitors' spatial behaviors in open plan exhibition settings. *Environment and Planning B: Planning and Design, 31*(3), 453-473.
Perera, L. D. H. (2014). Parents' attitudes towards science and their children's science achievement. *International Journal of Science Education, 36*(18), 3021-3041.
Perry, D. L. (2012). *What Makes Learning Fun? Principles for the Design of Instrinsically Moticating Museum Exhibits*. Walnet Creek, CA: AltaMira.
Persson, P.-E. (2015). *The impact of science centres.* Paper presented at the ECSITE Annual Conference, Trento, Italy.
Pezzullo, J. (2013). *Biostatistics for dummies*. Hoboken, NJ: John Wiley & Sons.
Phan, N. Q., Blome, C., Fritz, F., Gerss, J., Reich, A., Ebata, T., … Ständer, S. (2012). Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. *Acta dermato-venereologica, 92*(5), 502-507.
Phillips, C. M. L., & Beddoes, K. (2013). *Really Changing the Conversation: The Deficit Model and Public Under-standing of Engineering*. Paper presented at ASEE Annual Conference & Exposition, Atlanta, GA.
Piaget, J. (1968). *Six psychological studies.* New York, NY: Vintage Books.
Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human development, 15*(1), 1-12.
Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of gender on student's scientific reasoning ability: A case study in Thailand. *Procedia-Social and Behavioral Sciences, 116*, 486-491.
Plomin, R. (1986). *Development, genetics, and psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569.

Powers, N., & Kier, M. (2016). Meeting Diverse Needs Through Scientific Fluency: A synthesis of research-based recommendations that promote scientific literacy for all students. *Journal of Virginia Science Education, 10*(2), 18-26.

Pramanik, S., & Hussain, A. (2019). Text normalization using memory augmented neural networks. *Speech Communication, 109,* 15-23.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93-115.

Procheş, Ş. (2016). Descriptive statistics in research and teaching: are we losing the middle ground? *Quality & Quantity, 50*(5), 2165-2174.

Prokop, P., Tuncer, G., & Kvasničák, R. (2007). Short-term effects of field programme on students' knowledge and attitude toward biology: a Slovak experience. *Journal of Science Education and Technology, 16*(3), 247-255.

Raaijmakers, Q. A., Van Hoof, J., 't Hart, H., Verbogt, T. F., & Vollebergh, W. A. (2000). Adolescents' midpoint responses on Likert-type scale items: neutral or missing values? *International Journal of Public Opinion Research, 12,* 208-216.

Ramey-Gassert, L. (1997). Learning Science beyond the Classroom. *The Elementary School Journal, 97*(4), 433-450. doi:10.1086/461875

Randi Korn & Associates Inc. (2006). *Exhibit evaluation: Summative evaluation of Invention at Play*. Retrieved from http://www.exhibitfiles.org/dfile2/ReviewFinding/133/original/RK&A_Lemelson_Invention_small_summ_dist.pdf

Ravert, R. D., Gomez-Scott, J., & Donnellan, M. B. (2015). Equivalency of Paper Versus Tablet Computer Survey Data. *Educational Researcher, 44,* 308-310. doi:10.3102/0013189X15592845

Read, J., MacFarlane, S., & Casey, C. (2002, August 28-29). Endurability, engagement and expectations: Measuring children's fun. In *Proceedings of the International Workshop on 'Interaction Design and Children'* (pp. 189-198). Eindhoven, The Netherlands: Shaker Publishing.

Reilly, J. (2010). Practising gender equality in education. *Compare, 40*(4), 555-556.

Rendón-Macías, M. E., Villasís-Keeve, M. Á., & Miranda-Novales, M. G. (2016). Estadística descriptiva [Descriptive Statistics]. *Revista Alergia México* [Allery Magazine Mexico], *63*(4), 397-407.

Renner, N. O. (2011). *Multisensory sensemaking: Children's exploratory behavior has organizing structure at micro-and macro-scales.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society, Boston, MA.

Rennie, L. J., & Johnston, D. J. (2007). Visitors' perceptions of changes in their thinking about science and technology following a visit to science center. *Visitor Studies, 10*(2), 168-177.

Rennie, L. J., & Williams, G. F. (2002). Science Centers and Scientific Literacy: Promoting a Relationship with Science. *Science Education, 86,* 706-726. doi:10.1002/sce.10030

Renninger, K. A., & Hidi, S. (2002). Student interest and achievement: Developmental issues raised by a case study. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 173-195). San Diego, CA: Academic Press.

Research Centre for Museums and Galleries. (2003). *Measuring the outcomes and impact of learning in museums, archives and libraries*: *The Learning Impact Research Project End of Project Paper*. Retrieved from https://www.researchgate.net/publication/27244664_Measuring_the_outcomes_and_impact_of_learning_in_museums_archives_and_libraries_the_Learning_Impact_Research_Project_end_of_project_paper

Research New Zealand. (2018). *A survey of New Zealanders' perceptions of their national identity (2018)*. Retrieved from https://www.researchnz.com/special-reports.html#Indentity2018

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, 111*(12), 4403-4408.

Reynolds-Keefer, L., & Johnson, R. (2011). Is a picture is worth a thousand words? Creating effective questionnaires with pictures. *Practical Assessment, Research & Evaluation, 16*(8), 2.

Reynolds-Keefer, L., Johnson, R., Dickenson, T., & McFadden, L. (2009). Validity issues in the use of pictorial Likert scales. *Studies in Learning, Evaluation, Innovation and Development, 6*(3), 15-25.

Richardson, J. T. (2011). The analysis of 2× 2 contingency tables—Yet again. *Statistics in Medicine, 30*(8), 890-890.

Richter, J. G., Becker, A., Koch, T., Nixdorf, M., Willers, R., Monser, R., … Schneider, M. (2008). Self-assessments of patients via Tablet PC in routine patient care: comparison with standardised paper questionnaires. *Annals of the Rheumatic Diseases, 67*(12), 1739-1741.

Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociology of Education, 79*(3), 206-228.

Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher, 39*(9), 656-664.

Rio-Roberts, M. D. (2011). How I learned to conduct focus groups. *The Qualitative Report, 16*(1), 312-315.

Robinson, E. (1928). The behaviour of the museum visitor. In P. Bell, J. Fisher, & R. Loomis (Eds.), *Environmental Psychology*. Philadelphia, PA: WB Saunders.

Rodari, P., & Xanthoudaki, M. (2005). Beautiful Guides. The Value of explainers in science communication. *Journal of Science Communication, 4*(4), C01.

Roediger, H. L., III & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1155.

Roeser, S. (2012). Risk communication, public engagement, and climate change: a role for emotions. *Risk Analysis: An International Journal, 32*(6), 1033-1040.

Rogers, S. (2013). The pedagogization of play in early childhood education: A Bernsteinian perspective. In O. F. Lillemyr, S. Dockett, & B. Perry (Eds.), *Varied perspectives on play and learning* (pp. 159-174). Charlotte, NC: Information Age.

Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of Advanced Nursing, 53*(3), 304-310.

Roos, J. M. (2014). Measuring science or religion? A measurement analysis of the National Science Foundation sponsored science literacy scale 2006-2010. *Public Understanding of Science, 23*(7), 797-813. doi:10.1177/0963662512464318

Rosenthal, S. (2018). Motivations to seek science videos on YouTube: free-choice learning in a connected society. *International Journal of Science Education, Part B*, 8(1), 22-39.

Roshelle, J. (1995). Learning in interactive environments: Prior knowledge and new experience. In J. H. Falk & L. D. Dierking (Eds.), *Public institutions for personal learning: Establishing a research agenda*. (pp. 37). Washington, DC: American Association of Museums.

Rossiter, J. R. (2011). *Measurement for the Social Sciences: The C-OAR-SE Method and Why It Must Replace Psychometrics*. New York, NY: Springer-Verlag.

Rossman, G., & Rallis, S. (Eds.). (1998). *Learning in the Field: an Introduction to Qualitative Research*. Beverly Hills, CA: Sage.
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.
Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94-104.
Saldaña, J. (2015). *The Coding Manual for Qualitative Researchers*. Los Angeles,CA: SAGE Publications.
Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*, *36*(1), 43-53.
Salmi, H., Thuneberg, H., & Vainikainen, M.-P. (2015). Mathematical thinking skills, self-concept and learning outcomes of 12-year-olds visiting a Mathematics Science Centre Exhibition in Latvia and Sweden. *Journal of Science Communication, 14*(4), A03.
Sandifer, C. (1997). Time-Based Behaviors at an Interactive Science Museum: Exploring the Differences between Weekday / Weekend and Family / Nonfamily Visitors. *Science Education*, *81*, 689-701.
Sasao, T., Konomi, S. I., Arikawa, M., & Fujita, H. (2014). *Touch Survey: Comparison with Paper and Web Questionnaires*. Retrieved from https://arxiv.org/pdf/1405.4354.pdf
Sasson, I. (2014). The role of informal science centers in science education: attitudes, skills, and self-efficacy. *JOTSE, 4*(3), 167-180.
Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health, 2*(3), 223-228.
Schiele, B. (2008). Science museums and science centres. In M. Bucchi, & B. Trench (Eds.), *Handbook of Public Communication of Science and Technology* (pp. 41-54). New York, NY: Routledge.
Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology, 39*, 85-100.
Schunk, D., & Mullen, C. (2013). Motivation. In I. J. Hattie & E. M. Anderman (Eds.), *International Guide to Student Achievement* (pp. 67-69). New York, NY: Routledge.
Schwan, S., Bauer, D., Kampschulte, L., & Hampp, C. (2016). Representation Equals Presentation? *Journal of Media Psychology*, 29, 176-187.
Schwan, S., Grajal, A., & Lewalter, D. (2014). Understanding and engagement in places of science experience: Science museums, science centers, zoos, and aquariums. *Educational psychologist, 49*(2), 70-85.
Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, *54*(2), 93-105.
Scott, J., Brynin, M., & Smith, R. (1995). Interviewing children in the British household panel survey. In Stewart, I. & R. Vaitilingam (Eds.), *Advances in family research* (pp. 259-266). Essex, UK: University of Essex.
Scott, M., & Nenkov, G. (2013). “So Cute I Can Eat It Up”: Priming Effects of Cute Products on Indulgent Consumption. in S. Botti & A. Labroo (Eds.), *NA - Advances in Consumer Research Volume 41*. Duluth, MN: Association for Consumer Research.
Screven, C. G. (1992). Motivating visitors to read labels. *ILVS Review: A Journal of Visitor Behavior, 2*(2), 183-211.
Screven, C. G. (1999). Information design in informal settings: Museums and other public spaces. In R. Jacobson (Ed.), *Information Design* (pp. 131-192). Cambridge, MA: MIT Press.
Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. *BMJ, 351*:h4672.

Sefton-Green, J. (2012). *Learning at not-school: A review of study, theory, and advocacy for education in non-formal settings*. Cambridge, MA: MIT Press.

Şentürk, E., & Özdemir, Ö. F. (2014). The effect of science centres on students' attitudes towards science. *International Journal of Science Education, Part B, 4*(1), 1-24.

Sergeant, E. (2019). Epitools epidemiological calculators [Software]. Retrieved from http://epitools.ausvet.com.au

Serrell, B. (1997). Paying attention: The duration and allocation of visitors' time in museum exhibitions. *Curator: The Museum Journal, 40*(2), 108-125.

Serrell, B. (1998). *Paying Attention: Visitors and Museum Exhibitions*. Washington DC: American Association of Museums.

Serrell, B. (2006). *Judging Exhibitions: A Framework for Assessing Excellence*. Walnut Creek, CA: Left Coast Press.

Serrell, B. (2010). *Paying more attention to paying attention.* Retrieved from https://www.informalscience.org/news-views/paying-more-attention-paying-attention

Serrell, B. (2014). *A review of recommendations in exhibition summative evaluation reports*. Retrieved from BISE project resources webpage: https://www.visitorstudies.org/bise

Serrell, B. (2015). *Exhibit labels: An interpretive approach*. Lanham, MD: Rowman & Littlefield.

Shaby, N., Assaraf, O. B.-Z., & Tal, T. (2017). The particular aspects of science museum exhibits that encourage students’ engagement. *Journal of Science Education and Technology, 26*(3), 253-268.

Shaby, N., Assaraf, O. B.-Z., & Tishler, C. E. (2016). The goals of science museums in the eyes of museum pedagogical staff. *Learning Environments Research, 19*(3), 359-382.

Shah, K. N., Hofmann, M. R., Schwarzkopf, R., Pourmand, D., Bhatia, N. N., Rafijah, G., & Bederman, S. S. (2016). Patient-Reported Outcome Measures: How Do Digital Tablets Stack Up to Paper Forms? A Randomized, Controlled Study. *American Journal of Oorthopedics, 45*, E451-E457.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications, 9*(4787), 1-9. doi:10.1038/s41467-018-06930-7

Shemwell, J. T., Avargil, S., & Capps, D. K. (2015). Grappling with long-term learning in science: A qualitative study of teachers' views of developmentally oriented instruction. *Journal of Research in Science Teaching, 52*(8), 1163-1187.

Sherman, R. O. (2013). Imposter syndrome: When you feel like you’re faking it. *Am Nurse Today, 8*(5), 57-58.

Shettel, H. H., Butcher, M., Cotton, T., Northrup, J., & Slough, D. (1968). *Strategies for Determining Exhibit Effectiveness.* (Final Report). Washington DC: American Institutes for Research.

Shettel, H. H., & Reilly, P. C. (1966). Evaluation of existing criteria for judging the quality of science exhibits. *Educational Technology Research and Development, 14*(4), 479-488.

Shimamura, A. P., & Palmer, S. E. (Eds.). (2012). *Aesthetic science: Connecting Minds, Brains, and Experience*. New York, NY: Oxford University Press.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22-36.

Sidi, F., Panahy, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012). *Data quality: A survey of data quality dimensions.* Paper presented at the 2012 International Conference on Information Retrieval & Knowledge Management, Kuala Lumpur, Malaysia.

Siersdorfer, S., Minack, E., Deng, F., & Hare, J. (2010). *Analyzing and predicting sentiment of images on the social web.* Paper presented at the Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy.
Silvers, J. A., McRae, K., Gabrieli, J. D., Gross, J. J., Remy, K. A., & Ochsner, K. N. (2012). Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion, 12*(6), 1235.
Silvia, P. J. (2007). Knowledge-based assessment of expertise in the arts: Exploring aesthetic fluency. *Psychology of Aesthetics, Creativity, and the Arts, 1*(4), 247.
Silvia, P. J. (2013). Interested experts, confused novices: Art expertise and the knowledge emotions. *Empirical Studies of the Arts, 31*(1), 107-115.
Simon, N. (2010). *The participatory museum*. Santa Cruz, CA: Museum 2.0.
Simon, N. (2016). *The art of relevance*. Santa Cruz, CA: Museum 2.0.
Simon, R. M., & Farkas, G. (2008). Sex, class, and physical science educational attainment: Portions due to achievement versus recruitment. *Journal of Women and Minorities in Science and Engineering, 14*(3), 269-300.
Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*(1), 70.
Simpson, T. L. (2002). Dare I oppose constructivist theory?. *The Educational Forum, 66*(4), 347-354. doi:10.1080/00131720208984854
Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex roles, 50*(3-4), 241-252.
Smith, J. K. (2014). *The Museum Effect: How Museums, Libraries, and Cultural Institutions Educate and Civilize Society*. Lanham, MD: Rowman & Littlefield.
Smith, L. F., & Smith, J. K. (2006). The Nature and Growth of Aesthetic Fluency. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New directions in aesthetics, creativity, and the arts* (pp. 47-58). Amityville, NY: Baywood.
Smyth, J. M. (1998). Written emotional expression: effect sizes, outcome types, and moderating variables. *Journal of consulting and clinical psychology, 66*(1), 174.
Soper, D. S. (2019). A-priori Sample Size Calculator for Student t-Tests [Software]. Retrieved from http://www.danielsoper.com/statcalc
Stacey, K. (2010). Mathematical and scientific literacy around the world. *Journal of Science and Mathematics Education in Southeast Asia, 33*(1), 1-16.
Stange, M., Barry, A., Smyth, J., & Olson, K. (2018). Effects of smiley face scales on visual processing of satisfaction questions in web surveys. *Social Science Computer Review, 36*(6), 756-766.
Statistics New Zealand. (2012). *New Zealand women are having their first child at age 30.* Retrieved May 12, 2019 from http://archive.stats.govt.nz/browse_for_stats/population/mythbusters/first-baby-at-30.aspx
Statistics New Zealand. (2019). *2013 New Zealand census*. Retrieved May 9, 2019 from http://nzdotstat.stats.govt.nz/wbos/Index.aspx
Stocklmayer, S. M., & Bryant, C. (2012). Science and the Public—What should people know? *International Journal of Science Education, Part B, 2*(1), 81-101. doi:10.1080/09500693.2010.543186
Stocklmayer, S. M., & Gilbert, J. K. (2002). New experiences and old knowledge: towards a model for the personal awareness of science and technology. *International Journal of Science Education, 24*(8), 835-858.
Stodolsky, S. S., Salk, S., & Glaessner, B. (1991). Student views about learning math and social studies. *American Educational Research Journal, 28*(1), 89-116.

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581-593.

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth, 3*(1), e27.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103.

Struminskaya, B., Weyandt, K., & Bosnjak, M. (2015). The Effects of Questionnaire Completion Using Mobile Devices on Data Quality. Evidence from a Probability-based General Population Panel. *Methods, Data, Analyses, 9*, 32. doi:10.12758/mda.2015.014

Sturm, H., & Bogner, F. X. (2010). Learning at workstations in two different environments: A museum and a classroom. *Studies in Educational Evaluation, 36*(1-2), 14-19.

Suárez-Alvarez, J., Pedrosa, I., Lozano Fernández, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149-158. doi: 10.7334/psicothema2018.33

Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education, 5*(4), 541-542.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education, 4*(3), 279-282.

Sullivan, L. E. (2009). *The SAGE glossary of the social and behavioral sciences*. Los Angeles, CA: SAGE Publications.

Švec, V., Lawley, J., Nehyba, J., Svojanovský, P., Šíp, R., Minaříková, E., … Slavík, J. (2017). *Becoming a Teacher: The Dance Between Tacit and Explicit Knowledge*. Brno, Czech Republic: Masaryk University Press.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.

The Dodd-Walls Centre for Photonic and Quantum Technologies. (2016). *2015 Annual report*. Retrieved from https://www.otago.ac.nz/dodd-walls/otago691892.pdf

The Dodd-Walls Centre for Photonic and Quantum Technologies. (2017). *2016 Annual report*. Retrieved from https://www.otago.ac.nz/dodd-walls/otago660659.pdf

The Dodd-Walls Centre for Photonic and Quantum Technologies. (2018). *2017 Annual report*. Retrieved from https://www.otago.ac.nz/dodd-walls/otago686307.PDF.pdf

The White Oak Institute. (2016). *Museum Indicators of Impact and Performance (MIIP)*. Retrieved from http://www.whiteoakassoc.com/library.html

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44*(5), 423-432.

Thuneberg, H., & Salmi, H. (2018). To know or not to know: uncertainty is the answer. *Journal of Science Communication*, *17*(2), A01.

Tobias, S., & Duffy, T. M. (Eds.). (2009). *Constructivist Instruction: Success or failure?* New York, NY: Routledge.

Toepoel, V., & Funke, F. (2014). *Investigating response quality in mobile and desktop surveys: A comparison of radio buttons, visual analogue scales and slider scales.* Paper presented at the American Association of Public Opinion Research Conference. Anaheim, CA.

Tomczak, M., & Tomaczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1), 19-25.

Tsai, C. I., & Thomas, M. (2011). When does feeling of fluency matter? How abstract and concrete thinking influence fluency effects. *Psychological Science, 22*(3), 348-354.

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management, 24*(1), 53-62.

Tsuchihashi, M., Saito, M., Yamanoshita, M., Katsura, H., Myodo, O., & Maeyama, K. (1978). *United States Patent No. US4125888A*. Retrieved from https://patents.google.com/patent/US4125888A/en

Tudor, S. L. (2013). Formal-non-formal-informal in education. *Procedia-Social and Behavioral Sciences, 76*, 821-826.

United States Census Bureau. (2013). *Women in STEM occupations: 1970 to 2011*. United States Department of Commerce. Retrieved from https://www.census.gov/content/dam/Census/newsroom/releases/2013/cb13-162_stem_female.pdf

UNICEF. (2007). *A human rights-based approach to education for all: A framework for the realization of children's right to education and rights within education*. New York, NY: United Nations Children’s Fund.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational psychologist, 39*(2), 111-133.

van Gelder, M. M., Vlenterie, R., IntHout, J., Engelen, L. J., Vrieling, A., & van de Belt, T. H. (2018). Most response-inducing strategies do not increase participation in observational studies: a systematic review and meta-analysis. *Journal of Clinical Epidemiology, 99*, 1-13.

van Laerhoven, H., van der Zaag-Loonen, H., & Derkx, B. H. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica, 93*(6), 830-835.

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let’s learn from cows in the rain. *PLoS One, 8*(7), e68967.

Vaske, J. J., Jacobs, M. H., Sijtsma, M. T., & Beaman, J. (2011). Can weighting compensate for sampling issues in internet surveys? *Human Dimensions of Wildlife*, 16(3), 200-215.

Venville, G., Rennie, L. J., Hanbury, C., & Longnecker, N. (2013). Scientists reflect on why they chose to study science. *Research in science education, 43*(6), 2207-2233. doi:10.1007/s11165-013-9352-3

Villasís-Keever, M. Á., & Miranda-Novales, M. G. (2016). El protocolo de investigación IV: las variables de estudio [Research protocol IV: the study variables]. *Revista Alergia México* [Allergy Magazine Mexico], *63*(3), 303-310.

Vimala, R. (2010). *Gender issues in higher education*. Bangkok, Thailand: UNESCO Asia and Pacific Regional Bureau for Education.

Waddimba, A. C., Scribani, M., Nieves, M. A., Krupa, N., May, J. J., & Jenkins, P. (2016). Validation of single-item screening measures for provider burnout in a rural health care network. *Evaluation & the Health Professions, 39*(2), 215-225.

Wadgave, U., & Khairnar, M. R. (2016). Parametric tests for Likert scale: For and against. *Asian Journal of Psychiatry, 24*, 67.

Wagoner, B., & Jensen, E. (2010). Science learning at the zoo: evaluating children’s developing understanding of animals and their habitats. *Psychoology & Society, 3*(1), 65-76.

Wall, B. (2015). Living in a colourful world. *Junior Journal,* 51, 1-4. Retrieved from http://instructionalseries.tki.org.nz/Instructional-Series/Junior-Journal/Junior-Journal-51-Level-2-2015/Living-in-a-Colourful-World

Waller, B. M., Peirce, K., Mitchell, H., & Micheletta, J. (2012). Evidence of public engagement with science: Visitor learning at a zoo-housed primate research centre. *PLoS One, 7*(9). doi:10.1371/journal.pone.0044680

Wang, H., & Wang, S. (2010). Mining incomplete survey data through classification. *Knowledge and information systems, 24*(2), 221-233.

Wang, M.-T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction, 43*, 16-26.

Wang, W.-C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On known unknowns: Fluency and the neural mechanisms of illusory truth. *Journal of Cognitive Neuroscience, 28*(5), 739-746.

World Association of Zoos and Aquariums. (2018). *Annual Report 2017*. Retrieved from https://www.waza.org/wp-content/uploads/2019/02/2017-WAZA-Annual-Report_Final.pdf

Wehmeyer, M. L., Abery, B. H., Zhang, D., Ward, K., Willis, D., Hossain, W. A., … Calkins, C. (2011). Personal self-determination and moderating variables that impact efforts to promote self-determination. *Exceptionality, 19*(1), 19-30.

Weijters, B., Geuens, M., & Baumgartner, H. (2013). The effect of familiarity with the response category labels on item response to Likert scales. *Journal of Consumer Research, 40*(2), 368-381.

Weinburgh, M. H., & Steele, D. (2000). The modified attitudes toward science inventory: Developing an instrument to be used with fifth grade urban students. *Journal of Women and Minorities in Science Engineering, 6*(1), 87-98.

Weiss, S. M., Indurkhya, N., & Zhang, T. (2010). *Fundamentals of Predictive Text Mining*. London, UK: Springer-Verlag.

Wellington, J. (1990). Formal and informal learning in science: The role of the interactive science centres. *Physics education, 25*(5), 247.

Wells, T., Bailey, J. T., & Link, M. W. (2013). Filling the Void: Gaining a Better Understanding of Tablet-based Surveys. *Survey Practice, 6*, 1-4.

Wender, I. (2004). Relation of Technology, Science, Self-Concept, Interest, and Gender. *Journal of Technology Studies, 30*(3), 43-51.

Werquin, P. (2007). *Terms, concepts and models for analyzing the value of recognition programmes.* Paper presented at the Report to RNFIL: Third Meeting of National Representatives and International Organisations, Vienna, Austria.

Werquin, P. (2010). *Recognition of non-formal and informal learning: Country practices.* Paris, France: Organisation De Coopération Et De Développement Économiques. Retrieved from https://www.oecd.org/education/skills-beyond-school/44600408.pdf

Wieland, D. C. (2015). *Using cognitive task analysis to capture expert instruction in division of fractions* (Doctoral dissertation). Retrieved from USC Digital Library.

Wilcox, R. R. (2010). *Fundamentals of Modern Staistical Methods: Substantially Improving Power and Accuracy* New York, NY: Springer-Verlag.

Wilkins, J. L. (2000). Special issue article: preparing for the 21st century: the status of quantitative literacy in the United States. *School Science and Mathematics, 100*(8), 405-418.

Wilkins, J. L. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education, 72*(4), 331-346.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3), 1-13.

Williams, C., Stanisstreet, M., Spall, K., Boyes, E., & Dickson, D. (2003). Why aren't secondary students interested in physics? *Physics education, 38*(4), 324.

Williams, W. M., & Ceci, S. J. (2012). When scientists choose motherhood: a single factor goes a long way in explaining the dearth of women in math-intensive fields. How can we address it? *American Scientist, 100*(2), 138-145.

Wilson, P. M., Petticrew, M., Calnan, M., & Nazareth, I. (2010). Effects of a financial incentive on health researchers' response to an online survey: a randomized controlled trial. *Journal of Medical Internet Research*, *12*(2), e13.

Wineman, J. D., & Peponis, J. (2010). Constructing spatial meaning: Spatial affordances in museum design. *Environment and Behavior, 42*(1), 86-109.

Witcomb, A. (2006). Interactivity: thinking beyond. *A Companion to Museum Studies, 39,* 353-361.

Wittkower, D. E. (2012). On the origins of the cute as a dominant aesthetic category in digital culture. In T. W. Luke & J. Huninger (Eds.), *Putting Knowledge to Work & Letting Information Play* (pp. 167-175). Rotterdam, The Netherlands: Sense Publishers.

Wood, E., & Attfield, J. (2005). *Play, learning and the early childhood curriculum* (2nd ed.). London, UK: Paul Chapman.

Woods-McConney, A., Oliver, M. C., McConney, A., Maor, D., & Schibeci, R. (2013). Science engagement and literacy: A retrospective analysis for Indigenous and non-Indigenous students in Aotearoa New Zealand and Australia. *Research in science education, 43*(1), 233-252.

Worcester, R. M., & Burns, T. R. (1975). Statistical examination of relative precision of verbal scales. *Journal of the Market Research Society, 17*(3), 181-197.

Yalowitz, S. S., & Bronnenkant, K. (2009). Timing and tracking: Unlocking visitor behavior. *Visitor Studies, 12*(1), 47-64.

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ science of learning, 3,* 8.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Ppinion Quarterly, 75*(4), 709-747.

Zajacova, A., Lynch, S. M., & Espenshade, T. J. (2005). Self-efficacy, stress, and academic success in college. *Research in Higher Education, 46*(6), 677-706.

Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching, 45*(9), 1036-1058.

Zhang, H., Sproat, R., Ng, A. H., Stahlberg, F., Peng, X., Gorman, K., & Roark, B. (in press). Neural Models of Text Normalization for Speech Applications. *Computational linguistics*. Retrieved from https://www.mitpressjournals.org/doi/abs/10.1162/COLI_a_00349

Zimmerman, H. T., Reeve, S., & Bell, P. (2010). Family sense-making practices in science center conversations. *Science Education, 94*(3), 478-505.

Zubrowski, B. (1982). An aesthetic approach to the teaching of science. *Journal of Research in Science Teaching, 19*(5), 411-416. doi:10.1002/tea.3660190509

Appendices

Appendix A. Data Collection Methods

A.1 Surveys and visual counting

Four surveys were conducted in total, one in Discovery World and three in Tūhura. All of them were conducted on www.surveygizmo.com on iPad. In all cases, there was a paper version available in case someone preferred it. The paper version is the one shown here. There are some minimal differences between version, such as saying “circle” instead of “click on” to select some responses. Or having to write down the entry/exit times on the paper sheet, thing that was done automatically on iPad.

All surveys had aspects of knowledge, engagement and self-beliefs. The focus mentioned below is the main one for the particular survey. The pre and post versions are presented for all surveys.

Other documents related to the surveys, and the visual counting during the surveying period, are also presented.

A.1.1 Survey focused on knowledge in Discovery World

Pre-survey

Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Colour: ☐ Time of Entry: ☐

1. Tick one option for each statement that best shows what you think.	Disagree	Neither agree nor disagree	Agree	I don't know
I expect to enjoy Discovery World.	☐	☐	☐	☐
I think science museums are important.	☐	☐	☐	☐
Science is present in my life.	☐	☐	☐	☐
Scientists' work is important to me.	☐	☐	☐	☐
I could explain some science examples to my friends.	☐	☐	☐	☐

2. What three words would you use to describe Discovery World?

______________ ______________ ______________

3. Tick the one option that you think best fits the sentence.

What can be used to split white light into different colours?

☐ Mirror ☐ Prism ☐ Camera ☐ I don't know

If someone asks me: Can you see electromagnetic waves?, I will answer...

☐ Yes, all of them ☐ No, none of them ☐ I can see some of them ☐ I don't know

If someone asks me: Can atoms emit light?, I will answer...

☐ No, they can't ☐ Yes, but only in labs ☐ Yes, they can ☐ I don't know

I think electricity and magnetism are...

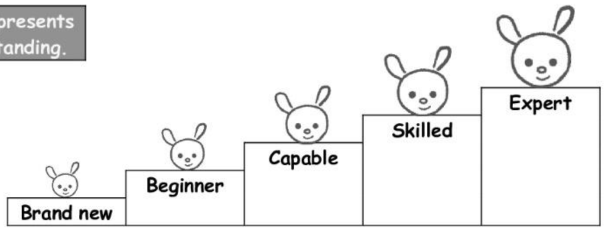
☐ Closely related ☐ Somewhat related ☐ Totally independent ☐ I don't know

4. Why did you come to Discovery World today? (Tick all that apply.)

☐ To see the Tropical Forest ☐ To play with the exhibits

☐ To see the Science Show ☐ Other:____________________

5. Circle the bunny that best represents yourself in the stair of science understanding.



6. About you

Your gender is: Male ☐ Female ☐ Other ☐

Your age is: 8 - 12 ☐ 13 - 18 ☐ 19 - 40 ☐ 41 plus ☐

You live in: Otago ☐ Rest of New Zealand ☐ Other ______________

Your ethnicity is: Māori / Pacific ☐ European ☐ Asian ☐ Other ______________

You came with: Family ☐ Friends ☐ School ☐ Other ______________

7. Including today, how many times have you come to Discovery World in your life?

☐ One ☐ Two or Three ☐ Four or more

If you have any suggestions or comments, this is your space.

Post-survey

Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Colour: ☐ Time of Leaving: ☐

1. Tick the option that best shows what you think.	Disagree	Neither agree nor disagree	Agree	I don't know
I enjoyed Discovery World.	☐	☐	☐	☐
I think science museums are important.	☐	☐	☐	☐
Science is present in my life.	☐	☐	☐	☐
Scientists' work is important to me.	☐	☐	☐	☐
I could explain some science examples to my friends.	☐	☐	☐	☐

2. What three words would you use to describe Discovery World?

__________ __________ __________

3. Tick the one option that you think best fits the sentence.

What can be used to split white light into different colours?

☐ Mirror ☐ Prism ☐ Camera ☐ I don't know

If someone asks me: Can you see electromagnetic waves?, I will answer...

☐ Yes, all of them ☐ No, none of them ☐ I can see some of them ☐ I don't know

If someone asks me: Can atoms emit light?, I will answer...

☐ No, they can't ☐ Yes, but only in labs ☐ Yes, they can ☐ I don't know

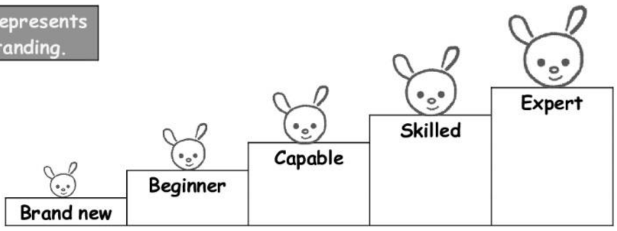
I think electricity and magnetism are...

☐ Closely related ☐ Somewhat related ☐ Totally independent ☐ I don't know

4. Who do you think Discovery World is suitable for? Tick all that apply.

☐ Children ☐ Teenagers ☐ Adults

5. Circle the bunny that best represents yourself in the stair of science understanding.



6. Complete the texts with your real thoughts.

It was cool learning about...	It would be cool if Discovery World had an exhibit about...

7. Did you do any of the following activities? (Tick all that apply.)

- ☐ Visit the Tropical Forest
- ☐ Play with the exhibits at Discovery World
- ☐ See the Science Show

If you have any suggestions or comments, this is your space.

A.1.2 Survey focused on knowledge in Tūhura

Pre-survey

Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Sticker: ☐ Time of Entry: ☐

Select the one option for each statement that best shows what you think.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	I don't know
I could explain some science examples to my friends.	☐	☐	☐	☐	☐	☐
I can understand new science ideas.	☐	☐	☐	☐	☐	☐
I am good at solving science problems that do not need math.	☐	☐	☐	☐	☐	☐
I am good at solving math problems.	☐	☐	☐	☐	☐	☐
I usually do well in science.	☐	☐	☐	☐	☐	☐
I learn science fast.	☐	☐	☐	☐	☐	☐
I have a good understanding of science.	☐	☐	☐	☐	☐	☐

What three words would you use to describe Discovery World?

__________ __________ __________

Tick the one option that you think best fits the sentence.

What can be used to split white light into different colours?

☐ Mirror ☐ Prism ☐ Camera ☐ I don't know

Can you see electromagnetic waves?

☐ Yes, all of them ☐ No, none of them ☐ I can see some of them ☐ I don't know

Can atoms give off light?

☐ No, they can't ☐ Yes, but only in labs ☐ Yes, they can ☐ I don't know

What colours can you combine to form white?

☐ Red+Green+Blue ☐ Cyan+Yellow+Magenta ☐ Red+Blue+Yellow ☐ I don't know

Does your body give off infrared radiation?

☐ No, never ☐ Yes, but only sometimes ☐ Yes, all the time ☐ I don't know

Which of the following things is necessary for an aurora to happen?

☐ An atmosphere ☐ Must be wintertime ☐ No moon in the sky ☐ I don't know

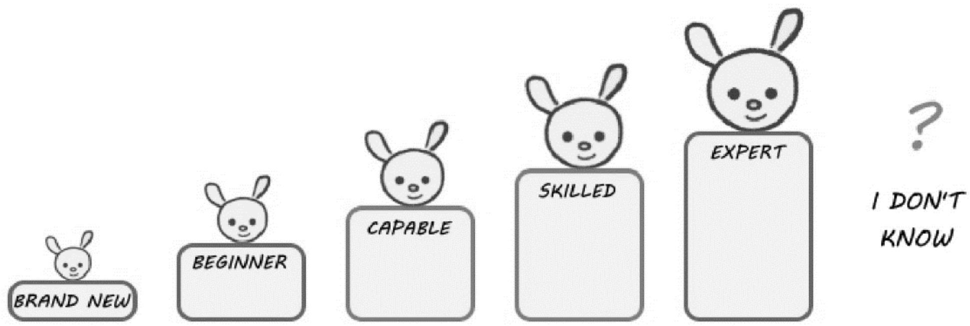
Why did you come to Discovery World today? (Tick all that apply.)

☐ See the Tropical Forest (butterfly house) ☐ Play with the exhibits

☐ Learn some science ☐ Go to the planetarium

☐ Other:_________________________

Click on the bunny that best represents yourself in the stair of science understanding.



About you

Your gender is: Male ☐ Female ☐ Other ☐

Your age is: ______ years old

You live in: Dunedin ☐ Rest of Otago ☐ Rest of New Zealand ☐ Other ____________________

Your ethnicity is (tick all that apply): Māori / Pacific ☐ European (Caucasian) ☐ Asian ☐ Other ____________________

You came with (tick all that apply): Family ☐ Friends ☐ Alone ☐ Other ____________________

Are you coming with young children today? Yes ☐ No ☐

Did you visit Discovery World before it closed?* Yes ☐ No ☐

*It was a science centre that used to be where Tūhura is now and closed in July 2017.

If you have any suggestions or comments, this is your space.

Post-survey

OTAGO museum   UNIVERSITY of OTAGO Te Whare Wānanga o Otāgo NEW ZEALAND   SCIENCE COMMUNICATION THE CENTRE FOR SCIENCE COMMUNICATION

Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Sticker: ☐ Time of Leaving: ☐

Including today, how many times have you come to Tūhura since it opened in December 2017? (Not including any visits to Discovery World)

☐ 1   ☐ 2   ☐ 3   ☐ 4 or more

Did you do any of the following activities? (Tick all that apply.)

☐ Visit the Tropical Forest (butterfly house)   ☐ Play with the exhibits

☐ Read the text at some exhibits   ☐ Went to the planetarium

☐ Other:___________________________

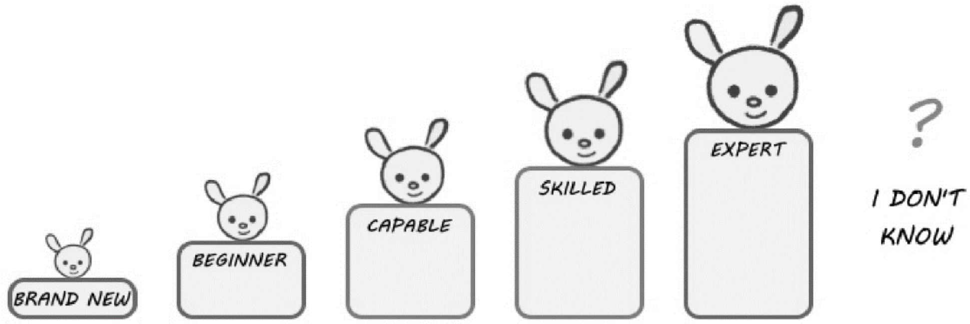
4. Who do you think Tūhura is suitable for? Tick all that apply.

☐ Children   ☐ Teenagers   ☐ Adults

What three words would you use to describe Discovery World?

______________________   ______________________   ______________________

Click on the bunny that best represents yourself in the stair of science understanding.



Select the one option for each statement that best shows what you think.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	I don't know
I could explain some science examples to my friends.	☐	☐	☐	☐	☐	☐
I can understand new science ideas.	☐	☐	☐	☐	☐	☐
I am good at solving science problems that do not need math.	☐	☐	☐	☐	☐	☐
I am good at solving math problems.	☐	☐	☐	☐	☐	☐
I usually do well in science.	☐	☐	☐	☐	☐	☐
I learn science fast.	☐	☐	☐	☐	☐	☐
I have a good understanding of science.	☐	☐	☐	☐	☐	☐

Complete the texts with your real thoughts.	
It was cool learning about...	It would be cool if Discovery World had an exhibit about...

Tick the one option that you think best fits the sentence.

What can be used to split white light into different colours?

☐ Mirror ☐ Prism ☐ Camera ☐ I don't know

Can you see electromagnetic waves?

☐ Yes, all of them ☐ No, none of them ☐ I can see some of them ☐ I don't know

Can atoms give off light?

☐ No, they can't ☐ Yes, but only in labs ☐ Yes, they can ☐ I don't know

What colours can you combine to form white?

☐ Red+Green+Blue ☐ Cyan+Yellow+Magenta ☐ Red+Blue+Yellow ☐ I don't know

Does your body give off infrared radiation?

- [ ] No, never
- [ ] Yes, but only sometimes
- [ ] Yes, all the time
- [ ] I don't know

Which of the following things is necessary for an aurora to happen?

- [ ] An atmosphete
- [ ] Must be wintertime
- [ ] No moon in the sky
- [ ] I don't know

Do you miss anything from Discovery World?*

*This question was only shown to visitors that selected that they visited Discovery World.

If you have any suggestions or comments, this is your space.

A.1.3 Survey focused on gender differences in Tūhura

Pre-survey

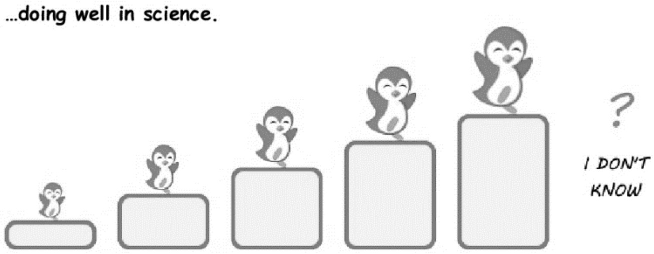
Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Sticker: ______ Time of Entry: ______

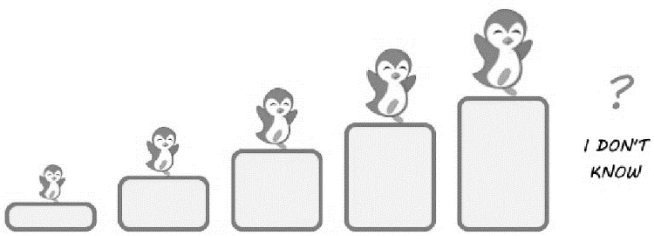
Select the one option for each statement that best shows what you think.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	I don't know
I would like to do some scientific research	☐	☐	☐	☐	☐	☐
I have a real desire to learn science	☐	☐	☐	☐	☐	☐
Science is fun	☐	☐	☐	☐	☐	☐
I would enjoy being a scientist	☐	☐	☐	☐	☐	☐
I enjoy learning science	☐	☐	☐	☐	☐	☐

Click on the penguin that best represents yourself in...

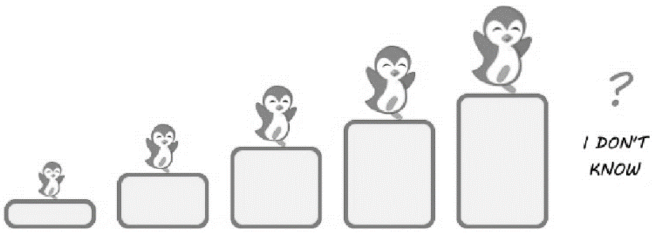
...doing well in science.



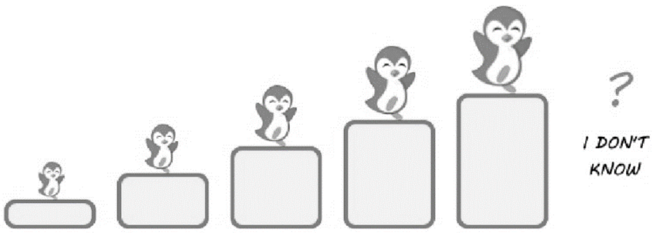
...learning science fast.



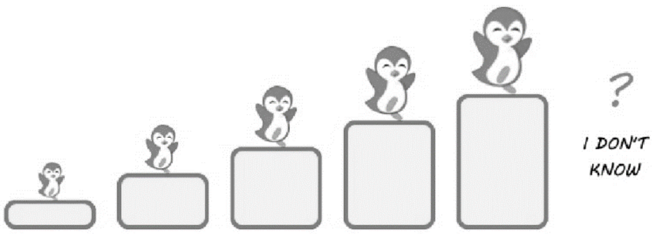
…science understanding.



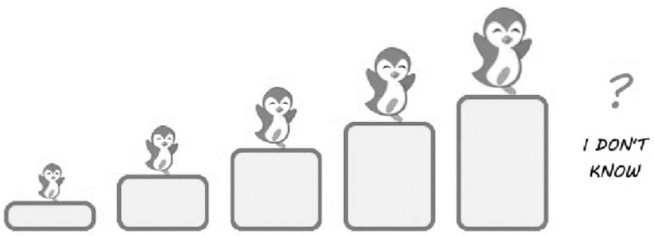
…ability to solve math problems.



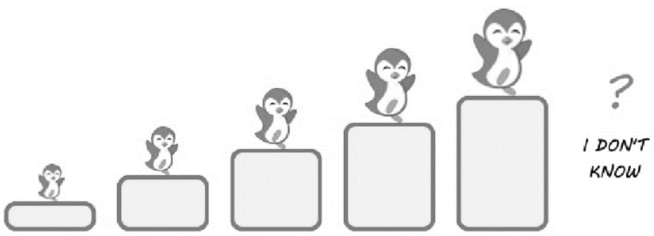
…confidence to understand new science ideas.



…ability to solve science problems that do not need math.

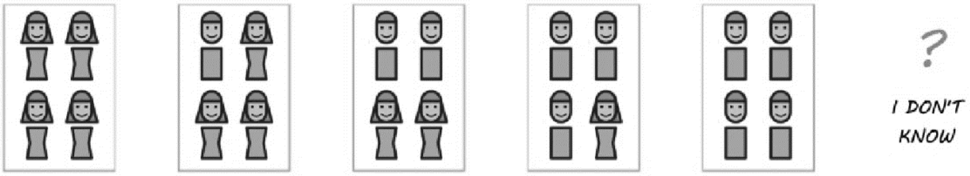


…confidence to explain some science examples to your friends.

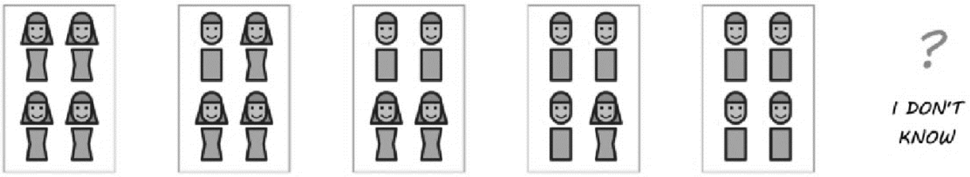


Click on the group that you think best answers each question.

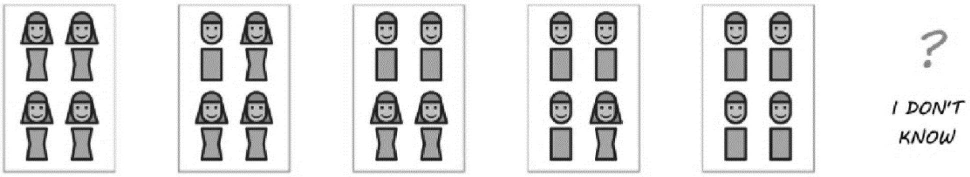
Which group of scientists is more likely to have made a very important discovery?



Which of the following groups of people would need to try harder to learn a new science idea?



Which of the following groups of people would likely be more confident in learning a new science idea?



Please indicate your level of familiarity with each of the following concepts:

Prism

- ☐ 0 – I have never heard of this
- ☐ 1 – I have heard of this but don't really know anything about it
- ☐ 2 – I have a vague idea of what this is
- ☐ 3 – I understand this when it is discussed
- ☐ 4 – I can talk intelligently about this

Plasma

- ☐ 0 – I have never heard of this
- ☐ 1 – I have heard of this but don't really know anything about it
- ☐ 2 – I have a vague idea of what this is
- ☐ 3 – I understand this when it is discussed
- ☐ 4 – I can talk intelligently about this

Infrared Light

- ☐ 0 – I have never heard of this
- ☐ 1 – I have heard of this but don't really know anything about it
- ☐ 2 – I have a vague idea of what this is
- ☐ 3 – I understand this when it is discussed
- ☐ 4 – I can talk intelligently about this

Magnetic Field

- ☐ 0 – I have never heard of this
- ☐ 1 – I have heard of this but don't really know anything about it
- ☐ 2 – I have a vague idea of what this is
- ☐ 3 – I understand this when it is discussed
- ☐ 4 – I can talk intelligently about this

Monochromatic Light

- ☐ 0 – I have never heard of this
- ☐ 1 – I have heard of this but don't really know anything about it
- ☐ 2 – I have a vague idea of what this is
- ☐ 3 – I understand this when it is discussed
- ☐ 4 – I can talk intelligently about this

Why did you come to Discovery World today? (Tick all that apply.)

- ☐ See the Tropical Forest (butterfly house)
- ☐ Play with the exhibits
- ☐ Learn some science
- ☐ Go to the planetarium
- ☐ Other:_________________________

About you

Your gender is:	Male ☐	Female ☐	Other ☐		
Your age is:	☐ years old				
You live in:	Dunedin ☐	Rest of Otago ☐	Rest of New Zealand ☐	Other ________________	
Your ethnicity is (tick all that apply):	Māori / Pacific ☐	European (Caucasian) ☐	Asian ☐	Other ________________	
You came with (tick all that apply):	Family ☐	Friends ☐	Alone ☐	Other ________________	
Are you coming with young children today?	Yes ☐	No ☐			

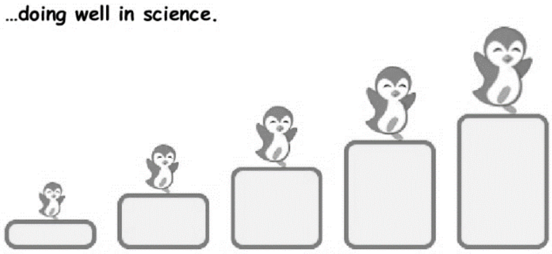
Post-survey

Thank you for your time. You are helping Discovery World with its continuous improvement.

Wristband Sticker: ______ Time of Leaving: ______

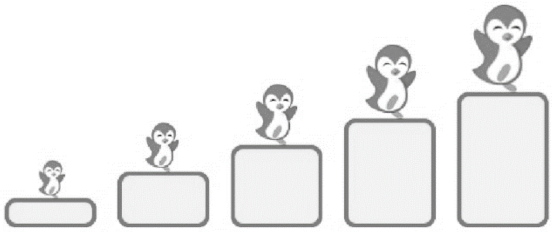
Click on the penguin that best represents yourself in...

...doing well in science.



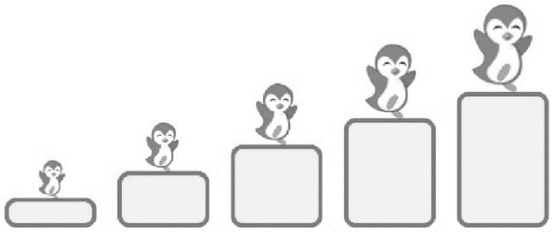
? I DON'T KNOW

...learning science fast.



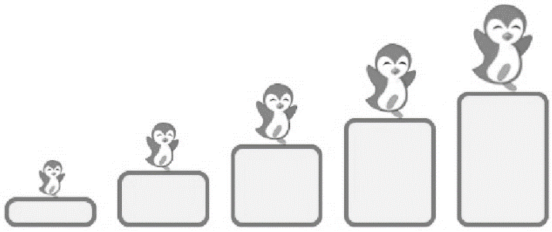
? I DON'T KNOW

...science understanding.



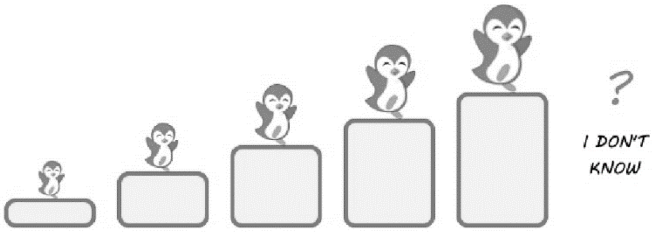
? I DON'T KNOW

...ability to solve math problems.

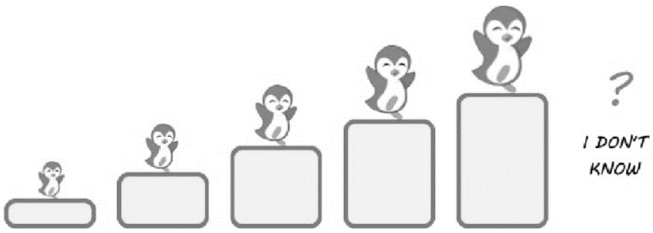


? I DON'T KNOW

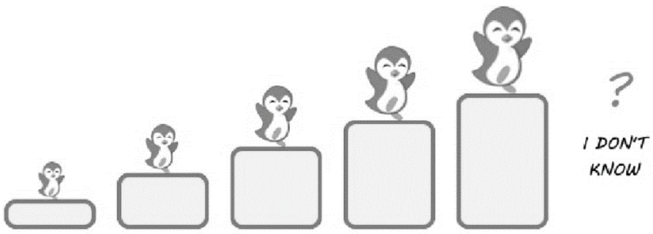
…confidence to understand new science ideas.



…ability to solve science problems that do not need math.

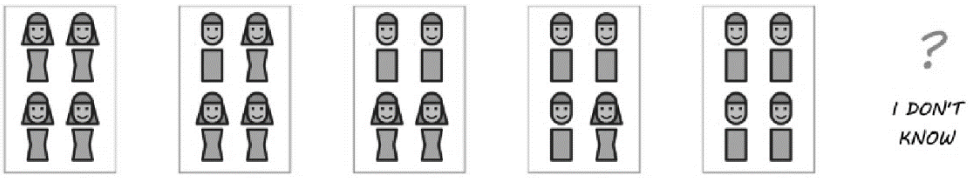


…confidence to explain some science examples to your friends.

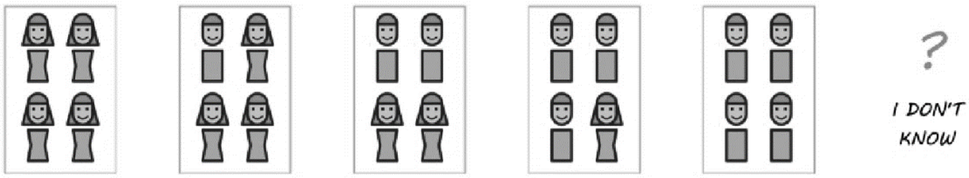


Click on the group that you think best answers each question.

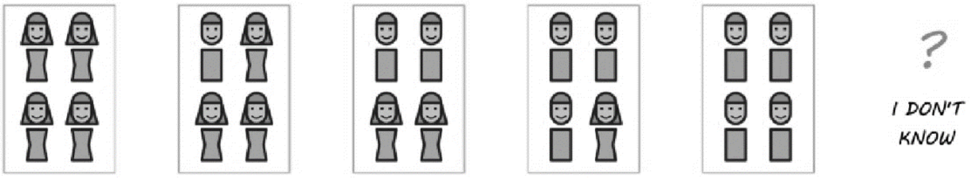
Which group of scientists is more likely to have made a very important discovery?



Which of the following groups of people would need to try harder to learn a new science idea?



Which of the following groups of people would likely be more confident in learning a new science idea?



Please indicate your level of familiarity with each of the following concepts:

Prism
- [ ] 0 – I have never heard of this
- [ ] 1 – I have heard of this but don't really know anything about it
- [ ] 2 – I have a vague idea of what this is
- [ ] 3 – I understand this when it is discussed
- [ ] 4 – I can talk intelligently about this

Plasma
- [ ] 0 – I have never heard of this
- [ ] 1 – I have heard of this but don't really know anything about it
- [ ] 2 – I have a vague idea of what this is
- [ ] 3 – I understand this when it is discussed
- [ ] 4 – I can talk intelligently about this

Infrared Light
- [ ] 0 – I have never heard of this
- [ ] 1 – I have heard of this but don't really know anything about it
- [ ] 2 – I have a vague idea of what this is
- [ ] 3 – I understand this when it is discussed
- [ ] 4 – I can talk intelligently about this

Magnetic Field
- [ ] 0 – I have never heard of this
- [ ] 1 – I have heard of this but don't really know anything about it
- [ ] 2 – I have a vague idea of what this is
- [ ] 3 – I understand this when it is discussed
- [ ] 4 – I can talk intelligently about this

Monochromatic Light
- [ ] 0 – I have never heard of this
- [ ] 1 – I have heard of this but don't really know anything about it
- [ ] 2 – I have a vague idea of what this is
- [ ] 3 – I understand this when it is discussed
- [ ] 4 – I can talk intelligently about this

Please indicate your level of familiarity with each of the following persons:

Albert Einstein

- ☐ 0 – I have never heard of this person
- ☐ 1 – I have heard of this person but don't really know anything about him
- ☐ 2 – I have a vague idea of who this person is
- ☐ 3 – I could understand a discussion about this person
- ☐ 4 – I can talk intelligently about this person

Ernest Rutherford

- ☐ 0 – I have never heard of this person
- ☐ 1 – I have heard of this person but don't really know anything about him
- ☐ 2 – I have a vague idea of who this person is
- ☐ 3 – I could understand a discussion about this person
- ☐ 4 – I can talk intelligently about this person

Isaac Newton

- ☐ 0 – I have never heard of this person
- ☐ 1 – I have heard of this person but don't really know anything about him
- ☐ 2 – I have a vague idea of who this person is
- ☐ 3 – I could understand a discussion about this person
- ☐ 4 – I can talk intelligently about this person

Marie Curie

- ☐ 0 – I have never heard of this person
- ☐ 1 – I have heard of this person but don't really know anything about her
- ☐ 2 – I have a vague idea of who this person is
- ☐ 3 – I could understand a discussion about this person
- ☐ 4 – I can talk intelligently about this person

Leonardo da Vinci

- ☐ 0 – I have never heard of this person
- ☐ 1 – I have heard of this person but don't really know anything about him
- ☐ 2 – I have a vague idea of who this person is
- ☐ 3 – I could understand a discussion about this person
- ☐ 4 – I can talk intelligently about this person

Including today, how many times have you come to Tūhura since it opened in December 2017? (Not including any visits to Discovery World, the science centre that used to be where Tūhura is now and that closed last year)

☐ 1 ☐ 2 ☐ 3 ☐ 4 or more

Did you do any of the following activities? (Tick all that apply.)

- ☐ Visit the Tropical Forest (butterfly house)
- ☐ Play with the exhibits
- ☐ Read the text at some exhibits
- ☐ Went to the planetarium
- ☐ Other:___________________________

Do you consider you learnt something at Tūhura's Exhibits that you did not know before? (including any previous visits).

☐ Yes ☐ No ☐ I haven't interacted with Tūhura's exhibits

Can you give an example of something you learnt?*

*This question was only displayed to visitors that selected that they considered they learnt something new.

If you visited any of the following attractions, please rate your experience in terms of Fun and Learning (leave blank those that you did not visit).

Tropical Forest (butterfly house)
Fun ★★★★★
Learning ★★★★★

Science Exhibits
Fun ★★★★★
Learning ★★★★★

Planetarium
Fun ★★★★★
Learning ★★★★★

If you have any suggestions or comments, either about Tūhura or the survey, this is your space.

A.1.4 Survey focused on self-beliefs in Tūhura

Pre-survey

Please, rate how confident you are that you can…	Not at all confident	A little confident	Somewhat confident	Moderately confident	Very confident	I don't know
find the position where a prism splits white light into different colours.	☐	☐	☐	☐	☐	☐
recognize plasma if you see it.	☐	☐	☐	☐	☐	☐
name characteristics of the infrared light.	☐	☐	☐	☐	☐	☐
test if a material is affected by a magnetic field.	☐	☐	☐	☐	☐	☐
recognize whether a colourful room is illuminated with monochromatic light.	☐	☐	☐	☐	☐	☐

Rank the following by order of importance for you, personally, to learn new scientific concepts. Tick 1 for the most important and 3 for the least:	1	2	3
Mathematics	☐	☐	☐
Explanations	☐	☐	☐
Experiments	☐	☐	☐

About you

Your gender is: Male ☐ Female ☐ Other ☐

Your age is: ____________

You live in: Dunedin ☐ Rest of Otago ☐ Rest of New Zealand ☐ Other ☐ ____________

Your ethnicity is: (tick all that apply) Asian ☐ Māori/Pacific ☐ European (Caucasian) ☐ Other ☐ ____________

Did you visit Discovery World* before it closed?: Yes ☐ No ☐ I'm not sure ☐

Why did you come to Tūhura* today? (Tick all that apply):

See the butterflies ☐ Play with the exhibits ☐ Learn some science ☐

Go to the planetarium ☐ Other ☐ ____________________

Wristband number: ____________ Date: ___/___/ 18 Entry time: ___:___

*The science centre you are visiting today was recently redeveloped and changed name. Until July 2017, it was known as Discovery World. From August to November it was closed for redevelopment. It opened back in December 2017, under the name of Tūhura.

v1

Wristband number: __________ Exit time: ___:___

Including today, how many times have you come to Tūhura since it opened in December 2017?:
(Not including any visits to Discovery World)

1 ☐ 2 ☐ 3 ☐ 4 or more ☐

Did you do any of the following activities? (Tick all that apply):

See the butterflies ☐ Play with the exhibits ☐ Read the text at some exhibits ☐

Go to the planetarium ☐ Other ☐ ______________________________

Please, rate how confident you are that you can…	Not at all confident	A little confident	Somewhat confident	Moderately confident	Very Confident	I don't know ?
find the position where a prism splits white light into different colours.	☐	☐	☐	☐	☐	☐
recognize plasma if you see it.	☐	☐	☐	☐	☐	☐
name characteristics of the infrared light.	☐	☐	☐	☐	☐	☐
test if a material is affected by a magnetic field.	☐	☐	☐	☐	☐	☐
recognize whether a colourful room is illuminated with monochromatic light.	☐	☐	☐	☐	☐	☐

Rank the following by order of importance for you, personally, to learn new scientific concepts. Tick 1 for the most important and 3 for the least:	1	2	3
Mathematics	☐	☐	☐
Explanations	☐	☐	☐
Experiments	☐	☐	☐

Please flip over the sheet of paper for last questions.

Tick the option that best shows what you think.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	I don't know
I discovered things that I didn't know.	☐	☐	☐	☐	☐	☐
I learnt more about things I already knew.	☐	☐	☐	☐	☐	☐
I remembered things I hadn't thought of for a while.	☐	☐	☐	☐	☐	☐
I shared some of my knowledge with other people.	☐	☐	☐	☐	☐	☐
I found the exhibition educational.	☐	☐	☐	☐	☐	☐
I got curious about finding out more about some things.	☐	☐	☐	☐	☐	☐

Have your say!

Do you have any ideas of how to improve any of the exhibits?
Share your ideas and comments with us!

Post-survey

Please, rate how confident you are that you can…	Not at all confident	A little confident	Somewhat confident	Moderately confident	Very Confident	I don't know
find the position where a prism splits white light into different colours.	☐	☐	☐	☐	☐	☐
recognize plasma if you see it.	☐	☐	☐	☐	☐	☐
name characteristics of the infrared light.	☐	☐	☐	☐	☐	☐
test if a material is affected by a magnetic field.	☐	☐	☐	☐	☐	☐
recognize whether a colourful room is illuminated with monochromatic light.	☐	☐	☐	☐	☐	☐

Rank the following by order of importance for you, personally, to learn new scientific concepts. Tick 1 for the most important and 3 for the least:

	1	2	3
Mathematics	☐	☐	☐
Explanations	☐	☐	☐
Experiments	☐	☐	☐

About you

Your gender is: Male ☐ Female ☐ Other ☐

Your age is: ____________

You live in: Dunedin ☐ Rest of Otago ☐ Rest of New Zealand ☐ Other ☐ ____________

Your ethnicity is: (tick all that apply) Asian ☐ Māori/Pacific ☐ European (Caucasian) ☐ Other ☐ ____________

Did you visit Discovery World* before it closed?: Yes ☐ No ☐ I'm not sure ☐

Why did you come to Tūhura* today? (Tick all that apply):

See the butterflies ☐ Play with the exhibits ☐ Learn some science ☐

Go to the planetarium ☐ Other ☐ ____________________

Wristband number: ____________ Date: ___/___/ 18 Entry time: ___:___

*The science centre you are visiting today was recently redeveloped and changed name. Until July 2017, it was known as Discovery World. From August to November it was closed for redevelopment. It opened back in December 2017, under the name of Tūhura.

v1

Wristband number: __________ Exit time: ___:___

Including today, how many times have you come to Tūhura since it opened in December 2017?:
(Not including any visits to Discovery World)

1 ☐ 2 ☐ 3 ☐ 4 or more ☐

Did you do any of the following activities? (Tick all that apply):

See the butterflies ☐ Play with the exhibits ☐ Read the text at some exhibits ☐

Go to the planetarium ☐ Other ☐ ____________________________

Please, rate how confident you are that you can…	Not at all confident	A little confident	Somewhat confident	Moderately confident	Very Confident	I don't know
find the position where a prism splits white light into different colours.	☐	☐	☐	☐	☐	☐
recognize plasma if you see it.	☐	☐	☐	☐	☐	☐
name characteristics of the infrared light.	☐	☐	☐	☐	☐	☐
test if a material is affected by a magnetic field.	☐	☐	☐	☐	☐	☐
recognize whether a colourful room is illuminated with monochromatic light.	☐	☐	☐	☐	☐	☐

Rank the following by order of importance for you, personally, to learn new scientific concepts. Tick 1 for the most important and 3 for the least:

	1	2	3
Mathematics	☐	☐	☐
Explanations	☐	☐	☐
Experiments	☐	☐	☐

Please flip over the sheet of paper for last questions.

Tick the option that best shows what you think.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	I don't know ?
I discovered things that I didn't know.	☐	☐	☐	☐	☐	☐
I learnt more about things I already knew.	☐	☐	☐	☐	☐	☐
I remembered things I hadn't thought of for a while.	☐	☐	☐	☐	☐	☐
I shared some of my knowledge with other people.	☐	☐	☐	☐	☐	☐
I found the exhibition educational.	☐	☐	☐	☐	☐	☐
I got curious about finding out more about some things.	☐	☐	☐	☐	☐	☐

Have your say!

Do you have any ideas of how to improve any of the exhibits?
Share your ideas and comments with us!

A.1.5 Information sheet (visitor survey)

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA
VISITOR SURVEY

INFORMATION SHEET FOR PARTICIPANTS or PARENTS / GUARDIANS

Thank you for your interest. Please read this information sheet before deciding whether or not to participate. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you and we thank you for considering our request.

Tūhura is the product of redeveloping Discovery World, the Otago Museum' science centre. The aim of this project to look at the outcomes of the redevelopment from the perspective of the visitors. We will provide feedback to those involved with Tūhura . This project is part of the PhD research of Daniel Solis.

Visitors are the cornerstone of this project. Visitors from 8 years old upwards are being asked to fill out a short survey before entering Discovery World and another before leaving and to wear a sticker on their visitor wristband so that we can match the pre and post surveys.

This research relates to your visit to Tūhura. Each survey will take about five minutes. You or your child may decide not to take part in the project or withdraw from participation in the project at any time without any disadvantage.

The survey will ask about you, but no identifying information will be collected. Anonymous quotes from suggestions, comments or open questions may be selected to be published verbatim. The only people who will have access to this project's raw data are the research team. Anonymous results will be summarized and provided to others, including those involved with Tūhura.

Data collected will be securely stored. Data obtained as a result of the research will be retained for at least five years in secure storage. The data derived from the research will, in most cases, be kept for much longer or possibly indefinitely. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve anonymity. A plain English summary of the findings will be available from the researchers' website at www.longneckerlab.net

Contact details for any questions about the project, now or in the future

Daniel Solis
PhD research student
Centre for Science Communication
021 0298 7337
daniel.solis@postgrad.otago.ac.nz

or

Nancy Longnecker
Professor of Science Communication
Centre for Science Communication
03 479 7885
nancy.longnecker@otago.ac.nz

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/062

A.1.6 Consent forms (visitor survey: Adults, Children, Parents/Guardians)

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA
VISITOR SURVEY

CONSENT FORM FOR PARTICIPANTS

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:

1. My participation in the project is entirely voluntary;

2. I am free to withdraw from the project at any time without any disadvantage;

3. The surveys may be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for at least five years;

4. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve my anonymity.

I agree to take part in this project.

....
(Signature of participant) (Date)

....
(Printed Name)

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/062

Reference Number: 17/062
25 May 2017

EVALUATION OF THE DISCOVERY WORLD REDEVELOPMENT
VISITOR SURVEYS

CONSENT FORM FOR CHILD PARTICIPANTS

I have been told about this study and understand what it is about. All my questions have been answered in a way that makes sense.

I know that:

1. My participation in this study is voluntary, which means that I do not have to take part if I don't want to and nothing will happen to me. I can stop taking part at any time and don't have to give a reason.
2. Anytime I want to stop, that's okay.
3. If I don't want to answer some of the questions, that's fine.
4. If I have any worries or if I have any other questions, then I can talk about these with the Daniel or the Discovery World science communicator.
5. The paper and computer file with my answers will only be seen by Daniel and the people he is working with. They will not use my name with any of my answers.
6. Daniel and his supervisors will write up the results from this study for their University work. The results may also be written up in journals and talked about at conferences. My name will not be on anything Daniel writes up about this study.

I agree to take part in the study.

....
Signed   Date

....
(Full name, please print)

Reference Number: 17/062
25 May 2017

EVALUATION OF TŪHURA
VISITOR SURVEY

CONSENT FORM FOR PARENTS/GUARDIANS

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:

1. My child's participation in the project is entirely voluntary;

2. I am free to withdraw my child from the project at any time without any disadvantage;

3. The surveys may be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for at least five years;

4. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve my child's anonymity.

I agree for my child to take part in this project.

....
(Signature of parent/guardian) (Date)

....
(Name of child)

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/062

A.1.7 Comments form

OTAGOmuseum

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

SCIENCE COMMUNICATION
THE CENTRE FOR SCIENCE COMMUNICATION

VISITOR COMMENTS

Day:		Time:	

Additional comments

A.1.8 Protocol for surveying visitors

OTAGO museum

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

SCIENCE COMMUNICATION
THE CENTRE FOR SCIENCE COMMUNICATION

EVALUATION OF THE DISCOVERY WORLD REDEVELOPMENT
PROTOCOL FOR SURVEY

Introduction:

- Introduce self as researcher / science communicator.
- Explain:
  a) The purpose of the survey,
  b) The nature of confidentiality,
  c) The use and purpose of the sticker on the wristband,
  d) That at the end of the visit there would be a second survey,
  e) That they do not have to answer every question and may withdraw at any time without prejudice,
  f) That they will receive a token as sign of gratitude.
- Ask if they have any additional questions or concerns before we commence.
- Ask the visitor to sign the consent form.
- At the conclusion of the survey, thank the participant and check whether they have any further questions or comments.

Notes:

- Focus only on children who are accompanied by a parent/guardian. Make the first contact with such parent/guardian.
- To know what visitor answered the pre-survey, a sticker with different colours and shapes are to be put to their visitor wristband. They select this shape and colour on the iPad.
- Have paper surveys in case someone is not comfortable with iPads. In that case, write down the time of entry and exit manually, as well as the shape and colour of the sticker.

A.1.9 Protocol for visual counting

EVALUATION OF TŪHURA
PROTOCOL: COUNTING PEOPLE

Write down the number of visitors that enter each time. Separate different counts by a slash. Try not to miss any. If there are any notable comments, write them down on the corresponding section. Use the back of the sheet if needed.

The code for special visitors is as follows:

F: Family
G: Group (3 or more visitors with no children)
C: Couple (2 visitors with no children, irrespective of the gender)
LV: Lone visitor (1 visitor)
B: Big group or family (5 or more members)
SG: School Group
ST: School Teacher
2G: Two generations
3G: Three generations

The code is preceded by a number, which specifies the number of visitors with the characteristics that follow. There is no need to specify the presence of children: Family implies the presence of children, while Group, Couple and Lone Visitor imply there are no children present (children are not allowed to go through on their own). If the number of generations is not specified, Family means 2G and Group, Couple and LV are 1G. The number of generations always goes at the end, to avoid confusions with the Group.

Examples:
1LV: a lone visitor
2C: 2 adults
3G2G: 3 adults in the same age range, but accompanied by someone of a different generation (noted down in a different section)
1F3G: 1 kid in a three-generation family
15SG: 1fifteen students in a school trip
3BF: three members of a family of 5 or more members in total

A.1.10 Visual counting sheet

EVALUATION OF THE DISCOVERY WORLD REDEVELOPMENT
SHEET FOR COUNTING PEOPLE

Day: | Sun | Mon | Tue | Wed | Thu | Fri | Sat | Period: | Schooldays | Holidays |

Date: ______ Time: ______ to ______

By Observations:

		Subtotal
2 minus		
	Non-Paying Children's Total	

2 – 7 / Male		
2 – 7 / Female		
8 – 12 / Male		
8 – 12 / Female		
13 – 18 / Male		
13 – 18 / Female		
	Children's Total	

19 – 40 / Male		
19 – 40 / Female		
41 plus / Male		
41 plus / Female		
	Adults' Total	

Comments:

A.1.11 Rejections sheet

SURVEY
REJECTIONS SHEET

Day:		Time:	

Rejections

No.	Possible cause / comments
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Additional comments

A.1.12 Non-monetary incentives

On a small piece of paper, one of the following facts was printed on the front and the institutional logos on the back. On the front, a present was taped. The presents were glow in the dark figures and fridge magnet butterflies (see Chapter 2).

<u>Glow in the Dark Facts:</u>

- Ostracods are small crustaceans that spit out light when they feel threatened!
- We can’t see it with our naked eye, but the human body glimmers very weakly while it metabolizes energy!
- 12,000 year old caves at Te Anau host a species of glowworm that will enlighten your trip to Milford Sound!
- Fireflies blink to flirt and to remind their enemies they don’t taste very good!
- Phosphorescent plankton lighting up the water has been spotted in several beaches around the world, including in New Zealand!
- Bioluminescence is the glow in the dark version of Nature. Some worms, snail shells and mushrooms can light up the night!
- The Waitomo Caves, in the North Island, are home of thousands of glowworms. This living blanket of fairy lights is unique to New Zealand!
- Your own glow in the dark present collects energy during the day and releases it at night in the form of light. It can because it contains *phosphors*!

<u>Magnetic Facts:</u>

- The Earth is like a big magnet. It has a magnetic north and a magnetic south, just like the magnet in your butterfly!
- Monarch butterflies have a compass in their antennae that is both sensitive to sunlight and magnetic fields. Try sticking your butterfly to your fridge!
- Most of the magnets we use are made of iron, like the one in your butterfly. Try sticking it to your fridge!
- Not all magnets are made of iron, your butterfly has one made of neodymium. Try sticking it to your fridge!

<u>Butterfly Facts:</u>

- Scientists thought butterflies were deaf until the first butterfly ears were identified in 1912!

- The eyes of butterflies are made of 6,000 lenses and can see ultraviolet light!
- Since butterflies are "cold-blooded", they can´t fly if the air temperature falls below 13 °C!
- Butterfly wings are transparent, the colours and patterns we see are made by the reflection of thousands of tiny scales covering them!

The following logos and website URL were on the back of each Facts Sheet.

www.longneckerlab.net

A.2 Interviews

Most interviews were conducted in two phases, before and after the redevelopment. When the interviewee was interviewed only once, questions from both phases were combined into a single interview. Only one interviewee dropped-out after the pre-redevelopment interview because she moved from the museum during the redevelopment.

The order of the documents is as follows:

1. Interview protocol
2. Information sheet
3. Consent forms
4. Science Communicators Discovery World interview
5. Decision Makers Discovery World interview
6. Decision Makers Tūhura interview
7. Science Communicators Tūhura interview
8. Maori curator complementary questions
9. Technician complementary questions

A.2.1 Interview protocol

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT
INTERVIEW PROTOCOL

- Introduce self as external researcher, independent of the redevelopment, that values sincerity and confidentiality.
- Double check that the participant understands the terms to which they earlier consented. Particularly:
  a) The purpose of the interview,
  b) The nature of confidentiality,
  c) That the interview will be recorded and
  d) That they do not have to answer a question they do not want to.
  e) That may withdraw at any time without prejudice.
- Ask if they have any additional questions or concerns before we commence.
- Start recording.
- Ask questions in semi-structured format, where they do not need to answer in the suggested order if a participant brings up a topic earlier or raises new questions.
- At the conclusion of the interview, thank the participant and check whether they have any further questions or comments.
- Stop recording.

A.2.2 Information sheet

Reference Number: D17/186
9 June 2017

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT / TŪHURA
INFORMATION SHEET FOR PARTICIPANTS

Please read this information sheet before deciding whether or not to participate. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you and we thank you for considering our request.

About participation in the project

This project examines perspectives on the Discovery World redevelopment. It is a part of research aimed at comparing the effectiveness of the old and new Discovery World in engaging visitors with science and facilitating learning. The research will provide feedback for the Otago Museum. We seek to interview people involved with the redevelopment of Discovery World and running of Tūhura.

This interview is a continuation of an interview conducted before Discovery World closed. You may withdraw from participation in the project at any time and without any disadvantage to yourself.

About the Data or Information collected and its use

Your interview will be audio-recorded and transcribed. A summary of your interview will be provided to you to check for accuracy. All the questions are directly related to the redevelopment of Discovery World and conceptions of science. The only people who will have access to this project's raw data are the researchers named below. Anonymized results will be summarized and provided to others, including Otago Museum staff.

The data collected will be securely stored and only the researchers will be able to gain access to it. Data obtained as a result of the research will be retained for at least 5 years in secure storage. Any personal information held on the participants will be destroyed at the completion of the research even though the data derived from the research will, in most cases, be kept for much longer or possibly indefinitely.

Results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but no identifying quote or material will be used without your express consent. A plain English summary will be available from the researchers' website: <www.longneckerlab.net>.

Contact details for any questions about the project, now or in the future:

Daniel Solis	or	*Nancy Longnecker*
PhD research student		Professor of Science Communication
Centre for Science Communication		Centre for Science Communication
021 029 87337		+(64) 3 479 7885
daniel.solis@postgrad.otago.ac.nz		nancy.longnecker@otago.ac.nz

This study has been approved by the Centre for Science Communication. However, if you have any concerns about the ethical conduct of the research you may contact the University of Otago Human Ethics Committee through the Human Ethics Committee Administrator (ph 03 479-8256), quoting the reference number D17/186. Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome.

1

A.2.3 Consent forms

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT / TŪHURA
CONSENT FORM FOR
PARTICIPANTS

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I understand that:

1. My participation is entirely voluntary;
2. I am free to withdraw at any time without any disadvantage;
3. I will be asked to answer questions in a semi structured interview. In the event that the line of questioning develops in such a way that I feel hesitant or uncomfortable I may decline to answer any particular question(s) and/or may withdraw without any disadvantage.
4. Any personal information (e.g. audio tapes) will be destroyed at the conclusion of the project, but raw data on which the results of the project depend on (e.g. interview transcripts) will be retained in secure storage for at least five years;
5. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt to preserve my anonymity will be made.

I agree to take part in this project.

....
(Signature of participant) (Date)

....
(Printed Name)

A.2.4 Science Communicators Discovery World interview

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT
PRE-REDEVELOPMENT (DW1) INTERVIEW
DISCOVERY WORLD SCIENCE COMMUNICATORS

Interview questions

Interviewee Background

1) Can you tell me briefly, what is your role at Discovery World?

2) Have you studied science?

   P.Q. Up to what level? What area?

Discovery World Redevelopment

3) Roughly, how many exhibits do you estimate you have interacted with at the current Discovery World? (Not necessarily the same day).

4) What are your most favourite exhibits in the current Discovery World?

5) What are the least favourite?

6) What do you expect from the redevelopment?

   P.Q. What do you expect from the new Discovery World?

7) Were you asked to provide input on how to redevelop Discovery World?

8) If redevelopment decisions were up to you, what are the chief modifications you would make to Discovery World?

Science

9) How would you define science outreach?

10) How would you define science?

11) What aspects of science are most important to get across in a science centre in general?

12) What do you know about the Dodd-Walls Centre?

Other

13) Are there any other thoughts you would like to share?

A.2.5 Decision Makers Discovery World interview

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT
PRE-REDEVELOPMENT (DW1) INTERVIEW
DESIGNERS AND OTAGO MUSEUM STAFF

Interview questions

Interviewee Background

1) Can you tell me briefly, what is your role at Gibson Group/ Otago Museum?

2) What is your role in the redevelopment of Discovery World?

3) Have you studied science?

   Probe: Up to what level? What area?

Discovery World Redevelopment

4) Have you visited the current version of Discovery World? If yes, roughly, how many exhibits do you estimate you have interacted with? (Not necessarily the same day).

5) What are your most favourite exhibits in the current Discovery World?

6) What are the least favourite?

7) Why do you think the redevelopment is needed?

8) What are the main goals of the redevelopment?

   Probe: What are the main objectives of the new Discovery World?

9) Are you going to evaluate the success of those goals or objectives? If so, how?

10) Are Science Communicators going to be trained to manage the new exhibits and concepts? If so, how and by whom?

Science

11) How would you define science outreach?

12) How would you define science?

13) What aspects of science are most important to get across in a science centre in general?

14) What do you know about the Dodd-Walls Centre?

Other

15) Are there any other thoughts you would like to share?

A.2.6 Decision Makers Tūhura interview

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT

POST-REDEVELOPMENT (**TŪHURA**) INTERVIEW

DECISION MAKERS

Interview questions

Discovery World Redevelopment

1) Has your role changed during the redevelopment, or now in the operation of Tūhura, with respect to the role you had in Discovery World?
   If yes: How?

2) Were Science Communicators trained in the use and conceptual understanding of Tūhura's exhibits?
   If yes 1: How?
   If yes 2: By whom?

3) Have you visited Tūhura?
   If yes: Roughly, how many exhibits do you estimate you have interacted with? Not necessarily the same day and not necessarily after the opening.

4) What are your most favourite exhibits in Tūhura and why?

5) What are your least favourite exhibits and why?

6) Which main goals or objectives of the redevelopment were fulfilled?

7) Which were not fulfilled?

8) Apart from this one, is the museum doing any evaluation of Tūhura that you know of?
   If yes: What is the evaluation saying so far?

9) What do you like about Tūhura?

10) What do you miss from Discovery World? If anything.

11) What challenges did you face during the redevelopment, or are you facing now that Tūhura is open?

Science

12) How would you define science communication?

13) How would you define science outreach?

14) How would you define science?

15) What aspects of science are most important to get across in a science centre in general?

16) What do you know about the Dodd-Walls Centre?

Other

17) Are there any other thoughts you would like to share?

A.2.7 Science Communicators Tūhura interview

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT

POST-REDEVELOPMENT (**TŪHURA**) INTERVIEW

SCIENCE COMMUNICATORS

Interview questions

Discovery World Redevelopment

1) Has your role changed during the redevelopment, or now in the operation of Tūhura, with respect to the role you had in Discovery World?
   If yes: How?

2) Were you trained in the use and conceptual understanding of Tūhura's exhibits?
   If yes 1: How?
   If yes 2: By whom?

3) Roughly, how many exhibits do you estimate you have interacted with at Tūhura? Not necessarily the same day and not necessarily after the opening.

4) What are your most favourite exhibits in Tūhura and why?

5) What are your least favourite exhibits and why?

6) Which of your personal expectations about Tūhura were fulfilled?
   Probe: Can you expand on that?
   Can you give me an example?

7) Which ones were not fulfilled?

8) What do you like about Tūhura?

9) What do you miss from Discovery World? If anything.

10) From your respective, have the demographics changed in Tūhura with respect to Discovery World?
   If yes. Probe: Can you expand in the comparison of how the redevelopment affected the attendance by gender and age group?

If no: What gender attends more?
If no: What age groups attend more?

11) Do you have an idea on how to enhance Tūhura to broaden demographics?
P. Q.: How would you make Tūhura more attractive to [mention gender that attended less, according to them, if any] or to [mention age groups they didn´t mention as those that attend more]?

Science

12) How would you define science communication?

13) How would you define science outreach?

14) How would you define science?

15) What aspects of science are most important to get across in a science centre in general?

16) What do you know about the Dodd-Walls Centre?

Other

17) Are there any other thoughts you would like to share?

A.2.8 Maori curator complementary questions

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT

POST-REDEVELOPMENT (**TŪHURA**) INTERVIEW

MAORI CURATOR

Māori knowledge

1) What term is more appropriate to describe what is conveyed in Tūhura, Mātauranga Māori, Mātauranga Pūtaiao, or some other term?

2) How would you describe Mātauranga Māori/Pūtaiao?

3) Why is the incorporation of Mātauranga Māori/Pūtaiao in Tūhura important?

4) Were there any particular challenges in incorporating Mātauranga Māori/Pūtaiao in Tūhura?

5) What is your personal opinion of the final result?

6) How has the incorporation of Mātauranga Māori/Pūtaiao in Tūhura been received by the public?

7) What does Mātauranga Māori/Pūtaiao bring to the science centre that it would be lacking otherwise?
   Probe: How does the inclusion of Mātauranga Māori/Pūtaiao complement the western science in Tūhura?

8) What do you expect Tūhura visitors to take away from their visit, regarding Mātauranga Māori/Pūtaiao?
   Follow up: Do you think they actually are taking that away?

9) What are the main differences and similarities between Mātauranga Māori/Pūtaiao and western knowledge?

10) Do you have any recommendations to other science centres in countries with a tradition in indigenous knowledge on how to incorporate indigenous knowledge into a science centre like Tūhura?

Interviewee Background

11) Can you tell me briefly, what is your role at the Otago Museum?*

12) Has your role changed during the redevelopment, or now in the operation of Tūhura, with respect to the role you had in Discovery World?
    If yes: How?

13) What was your role in the redevelopment of Discovery World?*

14) Have you studied science?*

    Probe: Up to what level? What area?

Science Centre Redevelopment

15) Did you visit Discovery World? If yes, roughly, how many exhibits do you estimate you have interacted with? (Not necessarily the same day)*

16) What were your favourite exhibits in Discovery World?*

17) What were the least favourite ones?*

18) Why do you think the redevelopment was needed?*

19) Were Science Communicators trained in the use and conceptual understanding of Tūhura's exhibits?
    If yes 1: How?
    If yes 2: By whom?

20) Have you visited Tūhura?
    If yes: Roughly, how many exhibits do you estimate you have interacted with? Not necessarily the same day and not necessarily after the opening.

21) What are your most favourite exhibits in Tūhura and why?

22) What are your least favourite exhibits and why?

23) What were the main goals of the redevelopment?*

    Probe: What are the main objectives of Tūhura?

24) Which main goals or objectives of the redevelopment were fulfilled?

25) Which ones were not fulfilled?

26) Apart from this one, is the museum doing any evaluation of Tūhura that you know of?
    If yes: What is the evaluation saying so far?

27) What do you like about Tūhura?

28) What do you miss from Discovery World? If anything.

29) What challenges did you face during the redevelopment, or are you facing now that Tūhura is open?

<u>Science</u>

30) How would you define science communication?

31) How would you define science outreach?

32) How would you define science?

33) What aspects of science are most important to get across in a science centre in general?

34) What do you know about the Dodd-Walls Centre?

<u>Other</u>

35) Are there any other thoughts you would like to share?

A.2.9 Technician complementary questions

STAFF VIEWS ON THE DISCOVERY WORLD REDEVELOPMENT

POST-REDEVELOPMENT (**TŪHURA**) INTERVIEW

TECHNICIAN

Technical perspective

1) How long have you been working at Tūhura?

2) Were you in charge of maintenance at Discovery World as well?

3) What are the specific challenges of keeping Tūhura's exhibits operational?

4) Do you have any preference among mechanical, computerized and gallery-type exhibits? Why?

5) After working for several months on Tūhura, what would you change about Tūhura?

6) If he worked at DW. How do Tūhura exhibits compare to Discovery World exhibits in terms of type and need of maintenance?

7) Do you have any recommendations for exhibit designers about how to make exhibits easier to maintain?

Interviewee Background

8) Can you tell me briefly, what is your role at the Otago Museum?*

9) Has your role changed during the redevelopment, or now in the operation of Tūhura, with respect to the role you had in Discovery World?
   If yes: How?

10) What was your role in the redevelopment of Discovery World?*

11) Have you studied science?*

    Probe: Up to what level? What area?

Science Centre Redevelopment

12) Did you visit Discovery World? If yes, roughly, how many exhibits do you estimate you have interacted with? (Not necessarily the same day)*

13) What were your favourite exhibits in Discovery World?*

14) What were the least favourite ones?*

15) Why do you think the redevelopment was needed?*

16) Were Science Communicators trained in the use and conceptual understanding of Tūhura's exhibits?
    If yes 1: How?
    If yes 2: By whom?

17) Have you visited Tūhura?
    If yes: Roughly, how many exhibits do you estimate you have interacted with? Not necessarily the same day and not necessarily after the opening.

18) What are your most favourite exhibits in Tūhura and why?

19) What are your least favourite exhibits and why?

20) What were the main goals of the redevelopment?*

    Probe: What are the main objectives of the new Discovery World?

21) Which main goals or objectives of the redevelopment were fulfilled?

22) Which ones were not fulfilled?

23) Apart from this one, is the museum doing any evaluation of Tūhura that you know of?
    If yes: What is the evaluation saying so far?

24) What do you like about Tūhura?

25) What do you miss from Discovery World? If anything.

26) What challenges did you face during the redevelopment, or are you facing now that Tūhura is open?

<u>Science</u>

27) How would you define science communication?

28) How would you define science outreach?

29) How would you define science?

30) What aspects of science are most important to get across in a science centre in general?

31) What do you know about the Dodd-Walls Centre?

<u>Other</u>

32) Are there any other thoughts you would like to share?

A.3 Focus Groups

The focus group pilot was conducted with a single group of children from 8 to 17 years old. The final version consisted of two groups, Children (8 to 12 years) and Adolescents (13 to 17 years).

The Protocol and the Example apply only for Discovery World. For Tūhura, they already knew how it worked and only a short reminder was needed.

The two extra questions mentioned in the protocol were asked only in Tūhura.

A.3.1 Protocol. It includes a SWOT example, guiding questions and the actual SWOT sheet for participants.

A.3.2 Information Sheets

A.3.3 Consent Forms

A.3.1 Protocol. It includes a SWOT example, guiding questions and the actual SWOT sheet for participants

EVALUATION OF THE DISCOVERY WORLD REDEVELOPMENT PROTOCOL: FOCUS GROUP v2

Conduct this instrument in redeveloped Discovery World at least six months after the first visit.

Participants:

- *Otago Museum's STEAM Team* members.
- *Fun Science with Amadeo (Kanavagh College) members.*
- *Children of Otago Museum Staff.*
- *2 groups. Between 6 and 10 participants per group. One group with children between 8 and 12 years old, and one group with adolescents between 13 and 17 years old.*

Specifications:

- The instrument is an adaptation of a SWOT (Strengths / Weaknesses / Opportunities / Threats) Analysis in a Focus Group.
- To facilitate the audio-recording and ordered discussion, "Talking Stick" technique is used (here called "Talking Butterfly", given that a butterfly figure will be attached to the stick). The stick will be a hand tripod with a cell phone mounted on it in audio-record mode.
- Given the time restrictions, if there are 8 or more participants, only suggestions approved by the majority will be discussed (thumbs up – approved, thumbs down – disproved, thumbs middle – null vote).
- Two audio-recording methods will be used: a recorder at the centre of the group and a cell phone on the Talking Butterfly. Also, notes will be taken during the Discovery World visit and the following discussion. SWOT results on the whiteboard may be photographed.
- Contrary to the original protocol, there will be no follow-up interview.
- Amadeo Enriquez and the researchers will be the only adults present, or at least, actively participating and speaking.
- During the Focus Group, participants will elaborate a cooperative SWOT chart, guided by the next related questions:

SWOT Topic	Question
Strengths	What is The Light Zone doing well? What makes The Light Zone to stand out positively in Discovery World? What is the most attractive of The Light Zone? What exhibits or topics you like the most in The Light Zone? What makes The Light Zone strong? …

Weaknesses	What is The Light Zone not doing well? What is not successful in The Light Zone? What is the least attractive of The Light Zone? What exhibits or topics you don't like in The Light Zone? What makes The Light Zone weak? …
Opportunities	What could The Light Zone do better? How would you improve The Light Zone? How could The Light Zone be more attractive to all ages? What exhibits or topics would you add to The Light Zone? What opportunity does The Light Zone have to improve? …
Threats	What could make visitors prefer not to play at The Light Zone? What could affect the number of visits to The Light Zone? What should The Light Zone avoid? What exhibits or topics would be better not to include in The Light Zone? What threats The Light Zone's success? …

Methodology:

1. Introduce oneself as a researcher and explain the objective of the instrument and what their participation consists in, including that they would be audio-recorded with two instruments, and Daniel will be taking notes during the visit and the discussion, but all results and quotes will be anonymized. Ask if they have questions. Where there are no more, give the participant consent forms to be signed.
2. Explain the SWOT (Strengths / Weaknesses / Opportunities / Threats) Analysis trough a collaborative example on the whiteboard.
3. Give the participants the sheet of paper with the SWOT sections and a pencil.
4. Ask participants to visit and play at Discovery World's Light Zone, but with a critic attitude and thinking on Strengths, Weaknesses, Opportunities and Threats.
5. Participants visit Discovery World's Light Zone for 30 minutes (they can also move around DW and play with the rest of the exhibits to compare).
6. After the visit, ask participants to sit down close to the whiteboard on a semi-circle.
7. Explain that Amadeo is going to work as a secretary, writing down on the whiteboard the in condensed sentence that is agreed in group. Daniel is going to work as a note-taker. Both will work as moderators, encouraging them to participate.
8. Explain that the participant with the Talking Butterfly has the right to speak and the rest should listen. If someone wants to participate, has to ask for the Talking Butterfly. Explain to them that this is for two reasons: to respect what others have to say, and to allow a clean recording.
9. Explain that once someone proposes an option, the rest will discuss (using the Talking Butterfly method) if it is accepted or not and with what modifications. They have to wait until the final proposal is written down on the whiteboard to propose a new one.
10. Turn on and put the recorder in the centre of the group. Turn on and put the cell phone-recorder on the Talking Butterfly.
11. Start with Strengths. Use tape to place the Strengths' guide questions close to the whiteboard.

12. Once that one has finished (try to finish in 10 mins), stop the recordings (save the files). Start new files and go on with the next SWOT section.
13. After all the four points in SWOT are covered, ask the following two questions:

Question 1. **What do you like about Tūhura´s exhibits on Light and Electromagnetism?**
Question 2. **What do you miss from Discovery World's Light Zone?**

14. Collect all the sheets of paper with the SWOT sections.

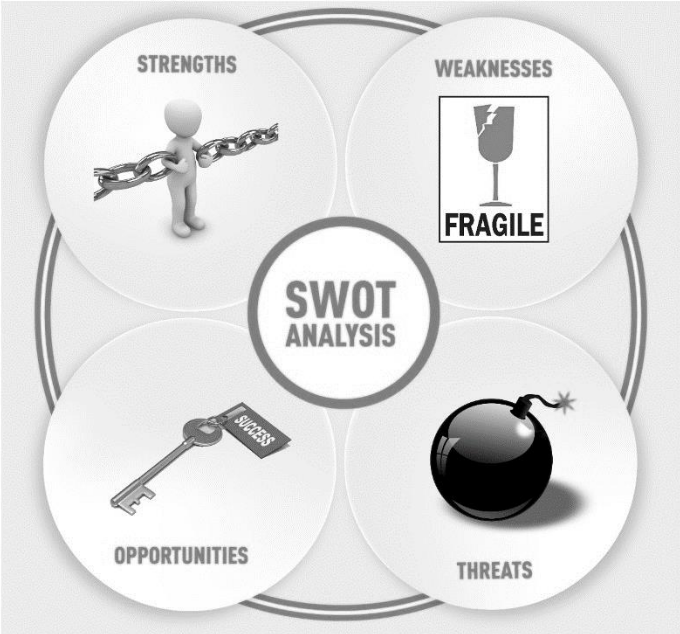


Figure 1. SWOT Analysis figure. Background image created by FreePix, superimposed images under Creative Commons rights.

SWOT Analysis Example: ACME (A Company that Makes Everything) Corporation is where Wile E. Coyote buys his stuff. Let's see an example of SWOT Analysis with this fictional company.

Strengths	Weaknesses	Opportunities	Threats
• What works well? • What are the advantages with respect to other companies where Willy could by his stuff?	• What disadvantages does it have? • What do other companies do better?	• What elements are not being exploited to full advantage? • What can be implemented to boost it?	• Which factors should be prevented? • What internal or external factors challenge its performance?

STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
➢ The market is the biggest possible because it makes everything. ➢ They have a fast delivery system.	➢ The lack of specialization results in lack of quality control, which is perceived by the consumer as products of low quality. ➢ Other companies offer warranty.	➢ Move from retailing to wholesale sales for the industry and third party companies. ➢ Absorbing subsidiary companies could make them specialists in some fields while still being a company for everything.	➢ Keep making everything implies a continuously growing infrastructure that could consume more than the available budget. ➢ The continuous failure of Willy can spread a bad name.

Strengths

- What is The Light Zone doing well?
- What makes The Light Zone to stand out positively in Discovery World?
- What is the most attractive of The Light Zone?
- What exhibits or topics you like the most in The Light Zone?
- What makes The Light Zone strong?
- …

Weaknesses

- What is The Light Zone not doing well?
- What is not successful in The Light Zone?
- What is the least attractive of The Light Zone?
- What exhibits or topics you don't like in The Light Zone?
- What makes The Light Zone weak?
- …

Opportunities

- What could The Light Zone do better?
- How would you improve The Light Zone?
- How could The Light Zone be more attractive to all ages?
- What exhibits or topics would you add to The Light Zone?
- What opportunity does The Light Zone have to improve?
- …

Threats

- What could make visitors prefer not to play at The Light Zone?
- What could affect the number of visits to The Light Zone?
- What should The Light Zone avoid?
- What exhibits or topics would be better not to include in The Light Zone?
- What threats The Light Zone's success?
- …

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA

DEMOGRAPHICS

About you

Your gender is: Male ☐ Female ☐ Other ☐

Your age is: ☐ years old

Your school year is ☐

You live in: Dunedin ☐ Rest of Otago ☐

Your ethnicity is (tick all that apply): Māori / Pacific ☐ European (Caucasian) ☐ Asian ☐ Other ____________

Did you participate in Discovery World Focus Group? Yes ☐ No ☐

SWOT sheet for participants (the actual sheet was printed 2-sided).

STRENGTHS	WEAKNESSES

OPPORTUNITIES	THREATS

A.3.2 Information Sheets

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA FOCUS GROUP
INFORMATION SHEET FOR PARENTS / GUARDIANS

Thank you for your interest in this project. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you and we thank you for considering our request.

Tūhura, the science centre at the Otago Museum, is the product of a major redevelopment last year. The aim of this project is to evaluate the redevelopment from the perspective of young people. The project will provide feedback to the parties involved with Tūhura and its redevelopment. This project forms part of PhD research by Daniel Solis.

We are seeking young students actively engaged with science, especially those that participated in the Focus Groups in Discovery World. Amadeo Enriquez, coordinator of *Fun Science with Amadeo*, and Craig Grant, Director of Science Engagement at the Otago Museum, have given consent and support.

All activities will be conducted in groups. First, the researcher will remind to the participants what the research consists in and what is a SWOT Analysis. Then, participants will make a group visit to Otago Museum's Tūhura where they will be free to interact with and critique the exhibits at the Light Zone to write down their opinion on sheets of paper. The group will then work together to discuss the Strengths, Weaknesses, Opportunities and Threats of the Light Zone. A couple of questions will complement the SWOT Analysis. The researcher and Amadeo Enriquez will be present at the discussion, taking notes and moderating it. The discussion will be audio recorded and transcribed for research purposes. The sheets of paper with the individual opinions will also be collected by the researcher.

Demographics will be asked, but no identifying information will be collected. If parents and child accept, photographs would be taken during these activities. Pictures would be taken by either Amadeo Enriquez or the researcher. A different consent form is included, but is independent, it's not necessary to agree to the photographs to participate in the Focus Groups.

The only people who will have access to this project's raw data are the research team. Anonymized results will be summarized and provided to others, including Otago Museum staff. Data will be securely stored so that only the researchers will be able to access it. Any personal information held on the participants will be destroyed at the completion of the research; the data derived from the research will be kept for at least five years in secure storage but possibly longer and even indefinitely.

Results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but no identifying quote or material will be used without your consent. A plain English summary of the findings will be available from the researchers' website at www.longneckerlab.net.

Contact details for any questions about the project, now or in the future:

Daniel Solis	or	*Nancy Longnecker*
PhD research student		Professor of Science Communication
Centre for Science Communication		Centre for Science Communication
021 0298 7337		03 479 7885
daniel.solis@postgrad.otago.ac.nz		nancy.longnecker@otago.ac.nz

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/06

A.3.3 Consent forms

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA FOCUS GROUP
CONSENT FORM FOR PARENTS/GUARDIANS

I have read the Information Sheet concerning this project and understand what it is about. My questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:

1. My child's participation in the project is entirely voluntary;
2. I am free to withdraw my child from the project at any time without any disadvantage;
3. This is the second part of the focus group, after the redevelopment, described in the information sheet;
4. This project involves developing a group Strengths, Weaknesses, Opportunities, Threats Analysis (SWOT) on Tūhura's the Light Zone during an audio recorded group discussion. In the event that the line of discussion develops in such a way that my child feels hesitant or uncomfortable, he/she may decline to participate and/or may withdraw from the project without disadvantage of any kind.
5. The paper results from the SWOT Analysis (SWOT table), as well as notes and audiotapes will be destroyed at the conclusion of the project, but any raw data on which the results of the project depend, will be retained in secure storage for at least five years;
6. The results of the project may be published and will be available in the University of Otago Library (Dunedin) but every attempt will be made to preserve my child's anonymity.

I agree for my child to take part in this project.

....
(Signature of parent/guardian) (Date)

....
(Name of child)

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/062

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA FOCUS GROUP

CONSENT FORM FOR CHILD PARTICIPANTS

I have been told about this study and understand what it is about. All my questions have been answered in a way that makes sense.

I know that:

1. Participation in this study is voluntary, which means that I do not have to take part if I don't want to and nothing will happen to me. I can also stop taking part at any time and don't have to give a reason.
2. Any time I want to stop, that's okay.
3. Daniel will audio record me so that *he* can remember what I say, but the recording will be erased after the study has ended.
4. If I don't want to participate in an activity, that's fine.
5. If I have any worries or if I have any other questions, then I can ask Daniel.
7. The SWOT table, notes, audio and computer files will only be seen or listened to by Daniel and the people he is working with.
8. Daniel and his supervisors will write up the results from this study for their University work. The results may also be written up in research journals and talked about at conferences. My name will not be on anything Daniel writes up about this study.

I agree to take part in the study.

....
Signed Date

....
Full name (please print)

Reference Number: 17/062
25 May 2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

EVALUATION OF TŪHURA PHOTOGRAPHS

CONSENT FORM FOR PARENTS/GUARDIANS

I have read the Information Sheet concerning this project and understand what it is about. In addition, the researcher has asked me for pictures of my child during the activity we previously agreed. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:

1. My child's appearance on the photographs is entirely voluntary;

2. I am free to withdraw my child from the photograph taking at any time without any disadvantage;

3. Photographs will be retained in secure storage for at least five years;

4. The photographs may be published on the thesis, articles, science outreach media, or presented in conferences. The thesis will be available in the University of Otago Library (Dunedin, New Zealand).

5. No names will be included with the photographs.

I agree for my child to take part in this project.

....

(Signature of parent/guardian) (Date)

....

(Name of child)

This study has been approved by the University of Otago Human Ethics Committee. If you have any concerns about the ethical conduct of the research you may contact the Committee through the Human Ethics Committee Administrator (ph +643 479 8256 or email gary.witte@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome. Reference number: 17/062

Reference Number: 17/062
25 May 2017

EVALUATION OF TŪHURA PHOTOGRAPHS

CONSENT FORM FOR CHILD PARTICIPANTS

I have been told about this study and understand what it is about. Also, the researcher asked to take pictures of me during the activity. All my questions have been answered in a way that makes sense.

I know that:

1. My appearance on the photographs is voluntary, which means that I do not have to accept if I don't want to and nothing will happen to me. If I don't want Daniel to take pictures of me, that's fine.

2. I can ask Daniel to stop taking pictures of me at any time if I no longer want to appear on them and don't have to give a reason.

4. If I have any worries or if I have any other questions, then I can talk about these with Daniel.

5. Daniel and his supervisors may publish or show the photos on their University work, journals, media o presented at conferences.

6. Photos not chosen to be published or presented at conferences will only be seen by Daniel and the people he is working with.

7. My name will not be shown with the photos.

I agree to take part in the study.

....
Signed Date

....
Full name (please print)

A.4 Other methods

The methods presented in this section where devised and, in some cases, piloted. But they were not used in their final version for different reasons, from technical issues to lack of time. Still, they are summarized here because of their potential to work in a different situation.

A.4.1 People Counter

Key Performance Indexes are a popular method to quantitatively assess the performance of museums, including science centres and their individual exhibitions. Time and attendance can be used in some of those indexes to measure visitors' behaviour.

The easiest method to record time and count visitors is observation, but it is also the most intrusive. A technological alternative is automatic visitors counting. Options for this include RFID, GPS localization, thermal and visual cameras, Bluetooth, etc. Despite these methods are in general unobtrusive and allow anonymous data collection, they tend to be expensive, especially when the system needs to be placed at several points.

A low-budget non-intrusive visitor counter was developed as part of this research. The system uses only readily available and affordable electronic parts. It consists of two devices, one emitting two pulsated infrared beams, and the other receiving them. Both of them are controlled by Arduino nano cards and work at the same peak wavelength and pulsation frequency.

The emitters and receivers are to be aligned and strategically placed at opposite sides of an exhibition's entrance. The beams should not interfere with each other and be at a suitable height to detect visitors. When a visitor obstructs any of the beams, a counter is activated until the receiver detects signal again. To avoid false positives, the distance between emitters/receivers needs to be short enough, so that a visitor detection is only considered valid if the second beam is interrupted while the first one has not presented continuity from the moment it was interrupted initially. The order in which the beams are cut determines the visitor's path direction. Every time a visitor is detected, direction, date and time are stored locally in a microSD.

Data coming from multiple devices covering all the entrances to the exhibition can be later read from the memory cards and merged to statistically infer crowdedness and average visit time. The main drawback of the system is the lack of reliability when several visitors cross the entrance simultaneously, but it is compensated by the low price.

The beta version was presented in the poster ‘Low-budget infrared directional device for counting visits to museum exhibitions’, presented at the New Zealand Institute of Physics Conference 2017, 10-12 July, in Dunedin, New Zealand.

Unfortunately, the final version broke down and there was no time to find out what caused the failure. The following images show the receiver and part of the code.

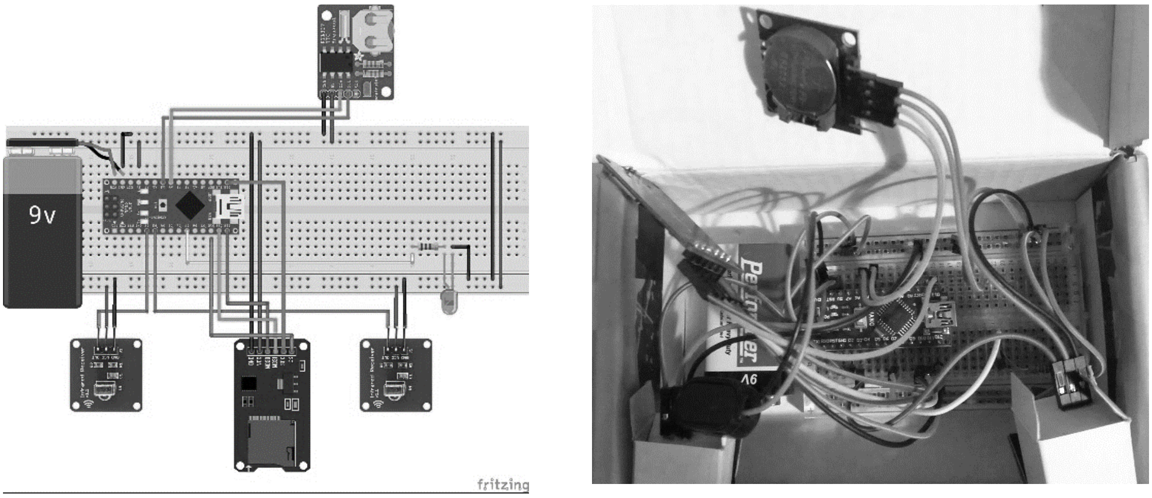
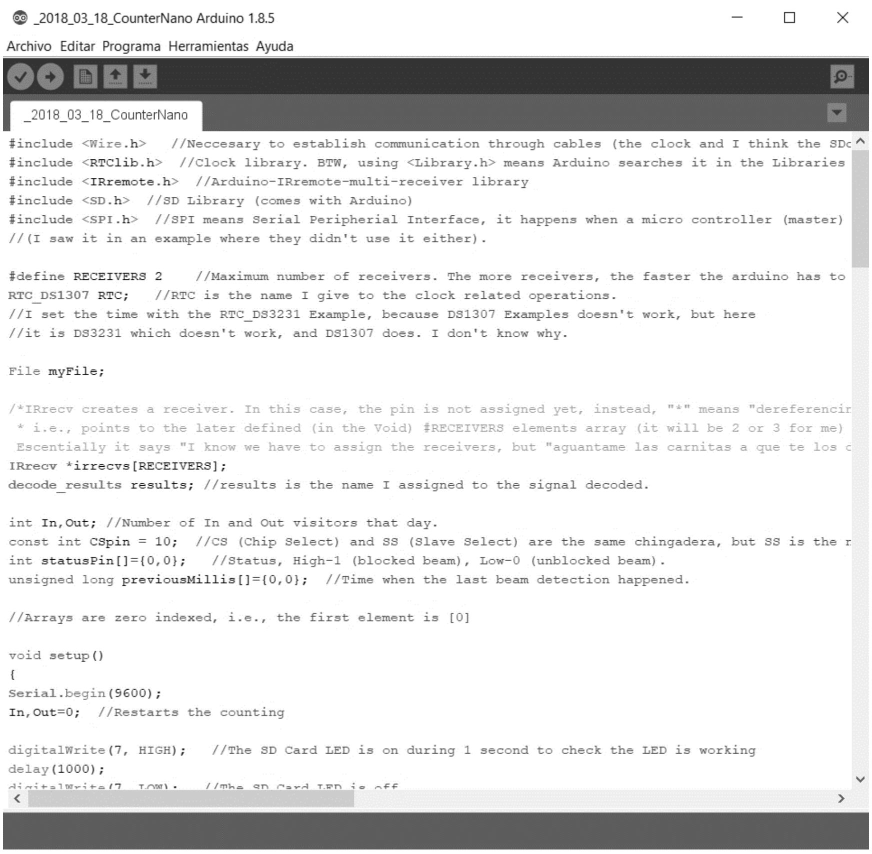


Figure. Receiver in design (left) and built (right).




```
#include <Wire.h>   //Neccesary to establish communication through cables (the clock and I think the SDc
#include <RTClib.h>  //Clock library. BTW, using <Library.h> means Arduino searches it in the Libraries
#include <IRremote.h>  //Arduino-IRremote-multi-receiver library
#include <SD.h>  //SD Library (comes with Arduino)
#include <SPI.h>  //SPI means Serial Peripherial Interface, it happens when a micro controller (master)
//(I saw it in an example where they didn't use it either).

#define RECEIVERS 2    //Maximum number of receivers. The more receivers, the faster the arduino has to
RTC_DS1307 RTC;   //RTC is the name I give to the clock related operations.
//I set the time with the RTC_DS3231 Example, because DS1307 Examples doesn't work, but here
//it is DS3231 which doesn't work, and DS1307 does. I don't know why.

File myFile;

/*IRrecv creates a receiver. In this case, the pin is not assigned yet, instead, "*" means "dereferencir
 * i.e., points to the later defined (in the Void) #RECEIVERS elements array (it will be 2 or 3 for me)
 Escentially it says "I know we have to assign the receivers, but "aguantame las carnitas a que te los c
IRrecv *irrecvs[RECEIVERS];
decode_results results; //results is the name I assigned to the signal decoded.

int In,Out; //Number of In and Out visitors that day.
const int CSpin = 10;  //CS (Chip Select) and SS (Slave Select) are the same chingadera, but SS is the r
int statusPin[]={0,0};   //Status, High-1 (blocked beam), Low-0 (unblocked beam).
unsigned long previousMillis[]={0,0};  //Time when the last beam detection happened.

//Arrays are zero indexed, i.e., the first element is [0]

void setup()
{
Serial.begin(9600);
In,Out=0;  //Restarts the counting

digitalWrite(7, HIGH);   //The SD Card LED is on during 1 second to check the LED is working
delay(1000);
digitalWrite(7, LOW);   //The SD Card LED is off
```

Screenshot of a section of the receiver’s code.

A.4.2 Mind Maps

Falk & Storksdieck (2005) conducted what they named a Personal Mind Mapping (PMM). In this method, visitors make a mind map before and after the visit, and clarify their answers in interviews. The advantage of mind maps (MM) is that they allow to measure the *extent* (the degree to which visitors can describe their understanding) and *breadth* (conceptual categories to describe the concept) of a concept through the use of words (Gondwe & Longnecker, 2015).

A variation of the PMM was used to assess learning in children that visited Discovery World during sleepovers. Children (usually coming from schools) spend a day and a night at Discovery World. Thanks to the museum's support, it was possible to access several of these children. However, the access to participants in this case was heavily limited by the conjunction of a number of factors. First, the sleepovers have a well-defined schedule that can hardly be modified. Second, to conduct the instrument, signed consent from the parents was needed, and the official path to reach them was long (researchers to museum staff to school representatives to school teachers to parents), resulting in most of the consents being granted in situ. Also, even when parents were usually cooperative, not all the children were accompanied by theirs and, therefore, could not participate. Third, several of the sleepovers were with too young children to participate. Fourth, the redevelopment was due to start in a very close date, after which, it was not possible to get more sleepover groups. Fifth, children sometimes like to stay awake late in sleepovers. One of the groups (n=6, Year 4) woke up very tired the second day, taking only 3 minutes to complete the post-mind map, as opposed to the 10 minutes they used the previous day. This group was not considered in the analysis.

As a result, despite all the efforts, only seven Year 11 ('Adolescents') and six Year 5 and 17 Year 4 (together referred as 'Children') were considered valid participants. Only 30 students from three groups of two schools were considered as valid participants. Data were collected from 26 June 2017 to 4 July 2017.

Interviews and follow-ups were removed from the PMM to reduce the time needed. Reductions of the PMM method with no interviews or follow-up have been conducted successfully in the past by Gondwe & Longnecker (2015).

Another modification was procedural. PMM implies returning the pre-visit mind map to the respondent after the visit, to add more concepts to the same MM. The problem is that any data collected with this method will produce positive results, there will always be more concepts in the post than in the pre (or at the very least, the same). However, memory is a complex and continuosly variable system. The idea was to see how the chosen concepts changed, not only if they could add more. Thus, participants received a blank sheet of paper in both times; after arrival to Discovery World (pre) and before departure (post). Typically, the pre was conducted at around 4pm and post 8am next day.

The topic for the mind map was ‘Light’. Children received an explanation of what a mind map is through a collaborative example where they, under the guide of the author, created a map around ‘Outer Space’.

The research questions this method was trying to answer were a) What light concepts are familiar to children? b) What light concepts are mentioned more often after a visit to Discovery World?

Children were allowed to work on the mind map as much as they wanted, but when most looked like not working anymore, they were told they could hand it in they finished. Those that looked that had finished tended to hand it in, while those still engaged continued working, not affected by those that left the room. The time needed to complete it varied from 10 to 15 minutes for children and eight to 13 minutes for adolescents.

The following images show typical mind maps.

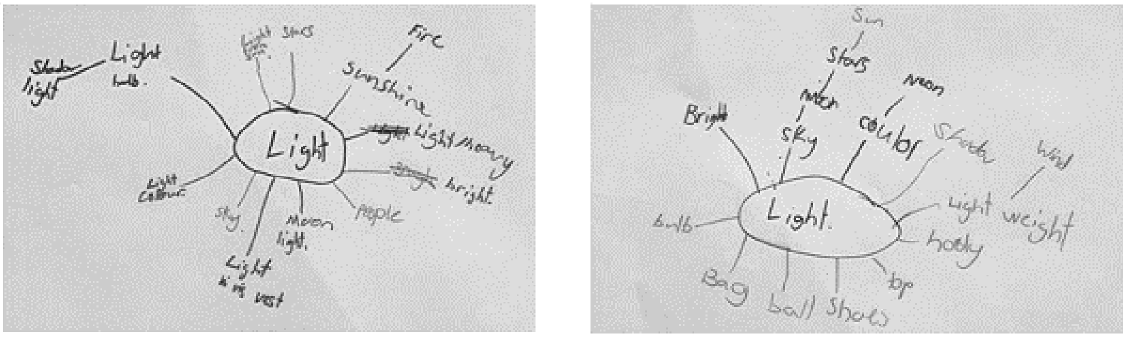


Figure. Example of Child’s mind maps, before (left) and after (right) the sleepover.

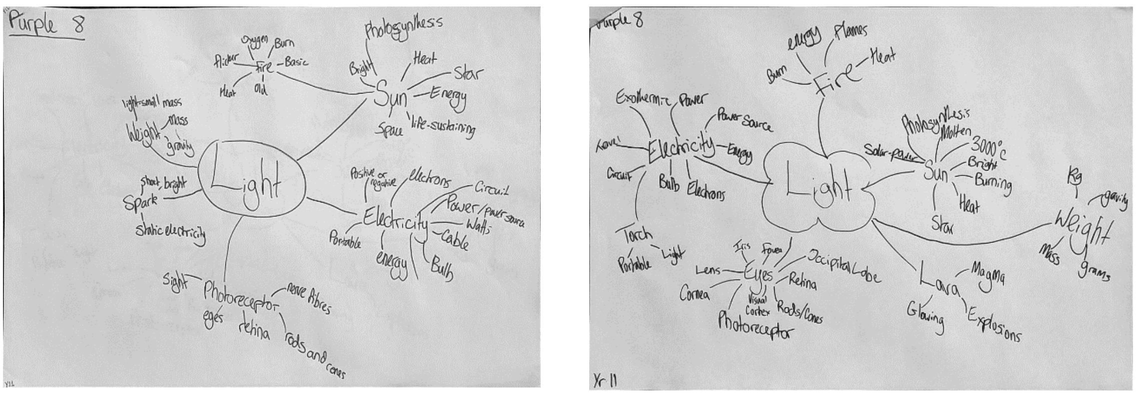


Figure. Example of Adolescent’s mind maps, before (left) and after (right) the sleepover.

Word clouds were made for Children and Adolescents with a similar pre-processing as described in Chapter 3 and Appendix D. This was the basic visual analysis. Children's word clouds do not show any big difference visually, being the main three words *bright*, *bun* and *lightbulb*.

Figure. Word clouds from Children's mind maps before (left) and after (right) the sleepover.

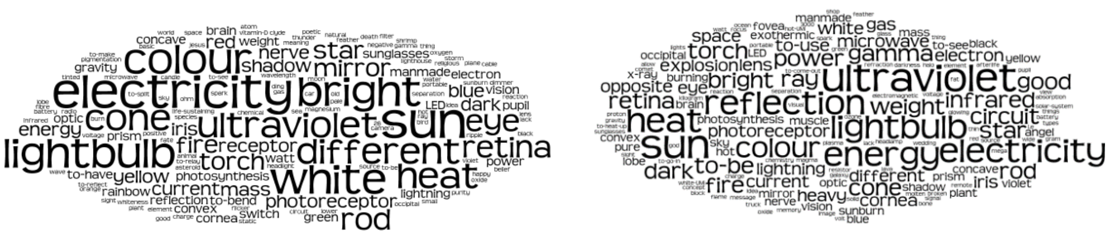


Figure. Word clouds from Adolescents' mind maps before (left) and after (right) the sleepover.

The total number of words used by Children (considering all the 23 valid responses) changed from 322 to 382 (increment of 19%). Adolescents changed from 283 to 368 (increment of 30%). The use of verbs increased from 10 to 14 (40%) in Adolescents, but decreased from 41 to 24 (-41%) in Children.

The following tables show the words that changed the most, taken as those with an increase or decrease of at least 4 or 3 in frequency depending of it corresponds to Children or Adolescents respectively. Words that the author considers fairly advanced are in bold. Only one of the seven words that show the most decrement is considered so (*receptor*). In those that increased its frequency, there are nine (*reflection*, *ultraviolet*, *laser*, *reflection*, *energy*, *infrared*, *circuit*, *ray*, *gamma*) out of 20.

Table
Words that decreased the most

Children			Adolescents		
	Pre	Post		Pre	Post
To see	11	7	White	7	3
Lighting	7	3	Bright	7	4
To do	5	1	Different	6	3
			Receptor	**3**	**0**

Table
Words that increased the most

Children			Adolescents		
	Pre	Post		Pre	Post
Star	4	11	**Reflection**	**2**	**6**
Sun	19	24	To be	1	4
Lights	4	8	Heavy	0	3
Different	5	9	Opposite	0	3
Torch	2	6	**Energy**	**3**	**6**
Reflection	**3**	**7**	**Infrared**	**1**	**4**
Rainbow	0	5	**Circuit**	**1**	**4**
Ultraviolet	**0**	**4**	Good	1	4
Laser	**0**	**4**	**Ray**	**1**	**4**
			Gamma	**1**	**4**
			To use	0	3

A Code Book was produced to analyse the mind maps through coding (not reproduced in full here). The Coding Book comprised three Coding Manuals:

A. Level. It simply referred to the level of the idea in the mind map, being 1 the minimum (0 corresponds to the central idea, *Light*).
B. Relation to Science (RS). It referred to what kind of environment it was more likely the idea was learnt in. The numbers were a quantitative measure of how close the idea was to be learnt in a scientific environment. Each idea was summed up to have a total score per mind map. 0. Not Related to Science (NRS, e.g. ‘Good’). 1. Learnt in Daily Life (DL, e.g. ‘Different colours’). 2. Learnt in Science Environment (SE, e.g. ‘Photosynthesis’).
C. Relation to Light (RL). It assessed if the idea was related to light as a luminous concept or not. Again, all ideas’ numbers were summed up to have a score for each mind map. 0. Not Related to Light (NRL, e.g. “heavy”). 1. Related to Light (RL, e.g. “Mirror”).
D. Conceptual Categories (CC). This one was the most comprehensive, as it included more categories. The previous ones were ordinal, but this one was nominal. The categories were: Natural Sources (NS, e.g. ‘Fire’). Artificial Sources (AS, e.g. ‘Laser’). Electromagnetic Spectrum (EMS, e.g. ‘UV’). Technology and Electricity (TE, ‘Current’). Energy and Chemistry (EC, e.g. ‘Heat’). Physics and Math (PM, e.g.

'Heavy'). Characteristics, Effects and Properties (CEP, e.g. 'Helps you see'). Biology (B, e.g. 'Eyes'). Unmeasurable Ideas (UMI, e.g. 'Poetry'). The full manual explained how to consider cases that apparently fit in more than one category.

The Code Book was a product of several iterations with a panel of experts. No inter-coder reliability was assessed as the method was not concluded. All analyses come from the author's coding.

Relation to Science, Relation to Light and Average Level were tested for significant changes with the Wilcoxon Signed Rank Test. Results are summarized in the following table. Something to note from this table is how different the mind maps are between Children and Adolescents. Children's mind maps are simple, almost single levelled with some relation to light and almost no relation to science. Adolescents' are more complex, with longer branches and more relation to light. Especially is to be noted the difference in terms of Relation to Science.

Table
Wilcoxon Signed Rank Test results for changes in level and relation to science and light

	Children (N=23)		Adolescents (N=7*)	
	Pre	Post	Pre	Post
Relation to Science	M=0.1, SD=0.4	M=0.7, SD=1.4	M=13.9, SD=7.5	M=17.7, SD=8.0
	Z=-2.410, p=0.016		Z=-1.016, p=0.310	
Relation to Light	M=10.2, SD=4.6	M=12.2, SD=4.9	M=21.4, SD=11.7	M=29.9, SD=7.3
	Z=-2.062, p=0.039		Z=-1.521, p=0.128	
Average Level	M=1.3, SD=0.3	M=1.3, SD=0.3	M=2.0, SD=0.3	M=2.1, SD=0.2
	Z=-1.490, p=0.136		Z=-0.734, p=0.463	

*Except for Average Level, where N=6, since one mind map was more visual with less connection and it was not possible to define branches.

The total number of ideas in the Conceptual Categories changed from 255 to 329 (29%) in Children and from 243 to 309 (27%) in Adolescents. These ideas were coded in the seven categories mentioned above. The following figures show the median changes for Children and Adolescents.

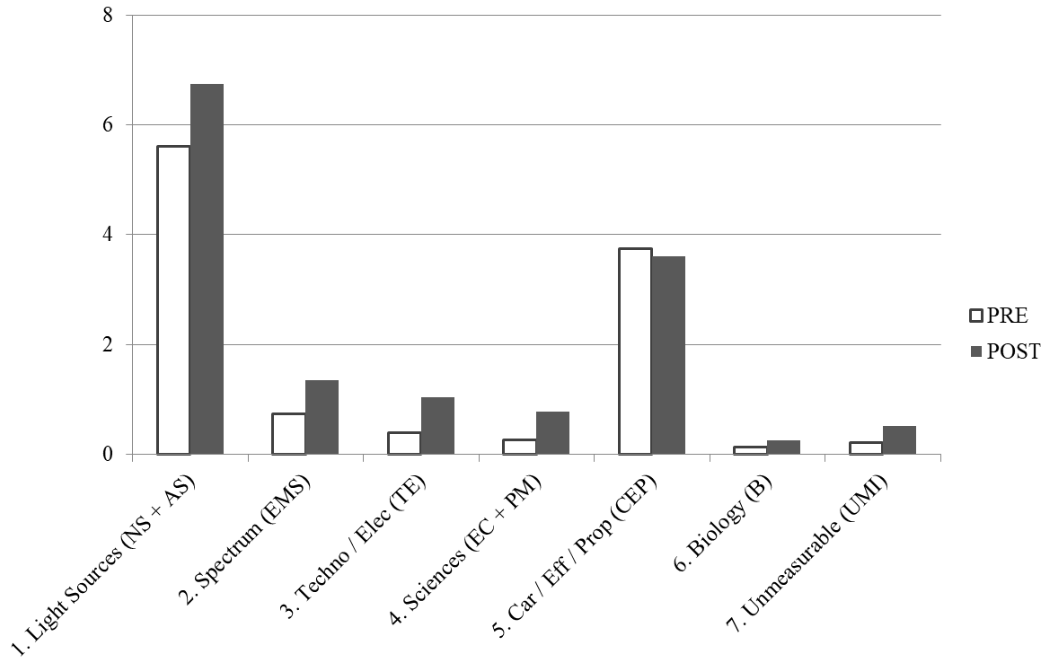


Figure. Mean number of Pre and Post ideas of Children.

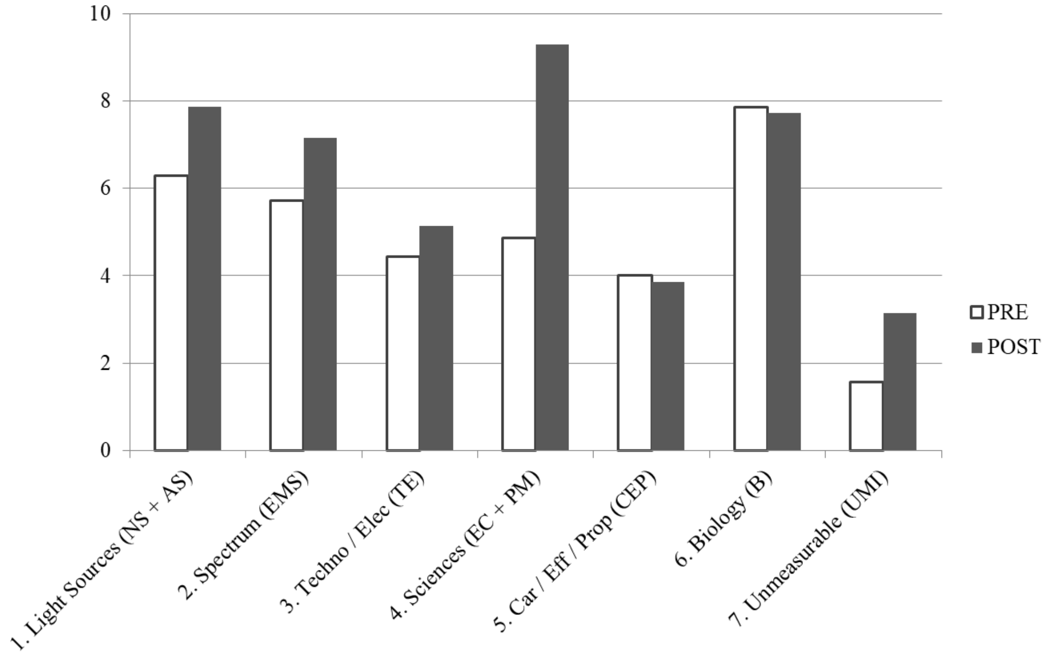


Figure. Mean number of Pre and Post ideas of Adolescents.

The Wilcoxon Signed Rank Test was conducted on all the categories to test for a significant change, as shown in the following table.

Table

Wilcoxon Signed Rank Test results for changes in categories

	Children (N=23)		**Adolescents (N=7)**	
	Pre	Post	Pre	Post
Light Sources	M=5.6, SD=3.7	M=6.7, SD=4.4	M=6.3, SD=3.8	M=7.9, SD=4.4
	Z=-2.048, p=0.041		Z=-1.511, p=0.131	
Electromagnetic Spectrum	M=0.7, SD=1.0	M=1.4, SD=1.4	M=5.7, SD=5.3	M=7.1, SD=5.8
	Z=-2.069, p=0.039		Z=-0.631, p=0.528	
Technology and Electricity	M=0.4, SD=0.7	M=1.0, SD=1.4	M=4.4, SD=2.6	M=5.1, SD=3.1
	Z=-2.215, p=0.027		Z=-0.742, p=0.528	
Sciences	M=0.3, SD=0.5	M=0.8, SD=1.1	M=4.9, SD=5.5	M=9.3, SD=6.6
	Z=-2.308, p=0.021		Z=-2.207, p=0.027	
CEP	M=3.7, SD=2.9	M=3.6, SD=1.9	M=4.0, SD=3.1	M=3.9, SD=2.1
	Z=-0.356, p=0.722		Z=-0.106, p=0.915	
Biology	M=0.1, SD=0.3	M=0.3, SD=0.4	M=7.9, SD=4.6	M=7.7, SD=4.9
	Z=-1.342, p=0.180		Z=-0.170, p=0.865	
Unmeasurable	M=0.2, SD=0.6	M=0.5, SD=0.9	M=1.6, SD=2.1	M=3.1, SD=3.5
	Z=-1.588, p=0.112		Z=-1.355, p=0.176	

Since post-maps seldom kept the same concepts mentioned in the pre-maps, it was decided not to continue. However, this may still be due to how the memory works. Memory is not a static machine that when accessed displays the same information. Memory is dynamic, any question we are asked today will elicit a different answer today and tomorrow. The appearance of significant changes in samples as small as N=23 and N=7 are possibly an indicative of actual significant changes.

A.4.3 Gender bias questions

To investigate gender bias in self-perceptions related to science, an instrument was developed with five independent exploratory items. The visitor had to move a continuous slider where a male and a female figures where at both ends.

Sliders have several advantages, like producing continuous data. Also, people prefer sliders over radio buttons for mobile devices as they are more interactive (Buskirk, Saunders, & Michaud, 2015) and respondents prefer them over Likert-type scales as an instrument to capture their true opinions (Cape, 2009; Toepoel & Funke, 2014). However, while all the claims above may hold for adults, slides turned out to be confusing for younger respondents and instrument, along with the data collected, were dismissed.

From the original five items, three were chosen to continue in a new discrete version where participants had to choose one of three figures of males and females with different proportions (50/50, all males, all females) or IDK. Due to the lack of significant findings, the options were later increased to five proportions plus IDK.

Table

Three items that were kept for the pilot in the discrete instrument about gender bias

CGB1. Which group of scientists is more likely to have made a very important discovery?
CGB2. Which of the following groups of people would need to try harder to learn a new science idea?
CGB3. Which of the following groups of people would likely be more confident in learning a new science idea?

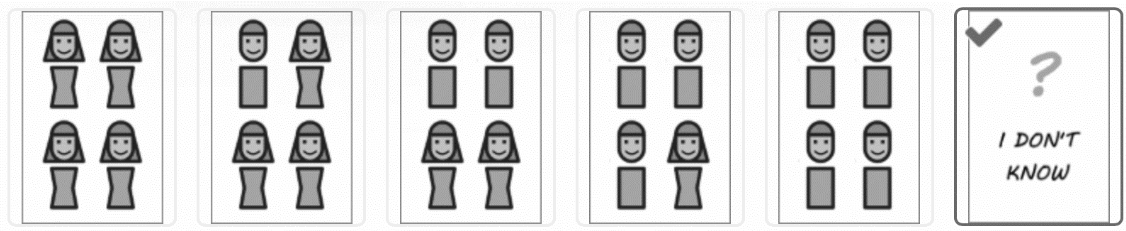


Figure. Anchors for questions about gender bias.

In total, 223 visitors filled out the 3-choices version and 163 the 5-choices version. In both cases, results were inconclusive. To facilitate statistics with the choices, the anchor with equal number of males and females (midpoint option) was identified as 0 (neutral), and the rest of the images were numbered with negative integers to the left (females, -1, -2) and positive integers to the right (males, +1, +2). IDK was considered as a missing value. Mann-Whitney Test was used to compare males and females one item at a time, before and after the visit for both set of responses 3-point items and 5-point items. For a Mann-Whitney test, reporting medians is usually the adequate, but means are reported for the sake of showing how small the differences are. Pre-post increments were tested with the Wilcoxon-Pratt test was conducted per gender in the three 5-point items. Results are shown in following tables.

Table

Gender differences by individual items in 3-point and 5-point versions

	N_{males}	$N_{females}$	U	p_{exact}	r	M_{male}	M_{female}
3-point items - Pre							
CGB1	78	109	3830.5	.098	0.1270	0.09	-0.02
CGB2	68	88	2960	.895	0.0108	0.12	0.15
CGB3	79	104	3850.5	.316	0.0737	0.01	-0.06
3-point items - Post							
CGB1	79	111	3734	0.010	0.1917	0.15	-0.01
CGB2	71	97	3430	1.000	0.0040	0.01	0.02
CGB3	77	108	3750	.137	0.1164	0.00	-0.11
5-point items - Pre							
CGB1	46	84	1830	.535	0.05639	0.17	0.02
CGB2	40	68	996.5	.011	0.2437	0.62	0.03
CGB3	48	80	1826	.504	0.0552	0.08	0.00
5-point items - Post							
CGB1	49	85	1666	0.020	0.1995	0.22	-0.20
CGB2	43	81	1193	0.001	0.2882	0.63	-0.09
CGB3	49	83	1682.5	0.040	0.1772	0.18	-0.16

Table
Pre/post differences by gender and by 5-point item

	N	Disc. Pairs	Z	p_{exact}	r
5-point items - Males					
CGB1	43	15	0.31363	0.7427	0.03382
CGB2	37	11	0.28171	0.9111	0.03275
CGB3	46	14	0.61319	0.5197	0.05418
5-point items - Females					
CGB1	79	22	1.3682	0.1647	0.10885
CGB2	64	22	0.45559	0.6558	0.04027
CGB3	77	23	1.4854	0.1554	0.01252

Results were inconclusive and the investigation was stopped. The reasons of the inconclusiveness may be several. In an attempt to not to bias the respondents through the images, the shapes were very similar and used the same neutral colours for both genders. Unfortunately, making the figures almost featureless confused some respondents, even adults, as exemplified by this short conversation between two middle-aged adults: "What do they represent?" (the man asking his wife while pointing at the gender items), "I think this is a woman" (replied the woman). One man directly asked the researcher because he couldn't distinguish, as all groups looked the same to him.

Most of the respondents didn't make any comments about these items. Those that did had a divided opinion. Some were positive, like a young female saying that "There were some interesting questions there, like the ones with male, female, male and female". And old man saying "This is interesting", and started talking about bias in science and unfair treatment for females.

But the opposite also happened. One time, a father and his young daughter were filling out the survey. When they reached the questions about gender, the parent took both iPads and put them on the desk while angrily saying "Sorry, not interested, this question totally pulled me out". Another time, a couple of young adults came through, he rejected, but she said yes, so the researcher gave the iPad to her, but he reacted grabbing it, almost taking it from her hands saying that they would fill it out together. When he/they reached the questions about gender, he literally tossed the iPad to the table and angrily yelled "I'm not going to answer something like that!" It is to be said that this behaviour only happened twice and in both cases the females (a child in one case, a partner in the other) didn't seem bothered by the questions.

It is possible that results were inconclusive because some respondents didn't distinguish the differences in the options and others were put off by the questions, or even some tried to give a socially acceptable answer, instead of their own opinion. "Survey respondents can give inaccurate answers to questions, especially those about normative behaviors” (Brenner & DeLamater, 2016, p. 349).

A.4.4 Ranking of science

“How you solve problems without maths” (T, M, 9) was what answered a visitor to the question “It was cool learning a bout…” The idea of this instrument was to detect of visitors were engaging with all the aspects of science or only with the spectacular part. Visitors were asked to rank Mathematics, Explanations and Experiments in order of importance to learn science before and after visiting Tūhura.

The three categories showed significant changes after the visit, as shown in the following table, but according to the Kendall’s Coefficient of Concordance, groups do not agree on the ranking in general. The only agreement case was 19+ females. It is more likely that changes are a product of randomness.

Table
Variation in ranking order due to the visit (N=164)

		1st	**2nd**	**3rd**	**Mean**	**DP**	**Wilcoxon-Pratt**
Mathematics	Pre	55	42	67	2.07	53	Z=10.4, p<.001
	Post	44	37	83	2.24		
Explanations	Pre	46	63	55	2.05	65	Z=10.8, p<.001
	Post	50	71	43	1.96		
Experiments	Pre	63	59	42	1.87	60	Z=10.7, p<.001
	Post	70	56	38	1.80		

DP stands for discordant pairs.

Table
Level of agreement among visitors by age and gender according to the Kendall's Coefficient of Concordance

			Pre		**Post**	
		n	**W**	**p**	**W**	**p**
8-18	Male	46	0.081	0.025	0.063	0.057
	Female	53	0.032	0.197	0.017	0.417
19+	Male	29	0.008	0.837	0.051	0.239
	Female	65	0.193	<0.001	0.290	<0.001

A.4.5 Marker Exhibit

An idea to assess the impact of the science centre was to have a marker exhibit, developed specifically to measure changes in understanding on light.

To understand the concept of colour temperature, an object would be illuminated with a lamp. The user could choose what light to use from 1000 K to 10,000 K. With the help of a spectrometer, it would be explained why a low temperature lamp (2700 K) can help us move in the streets at night, and still allow us to see the stars. And why the typical high temperature bulb hides the stars.

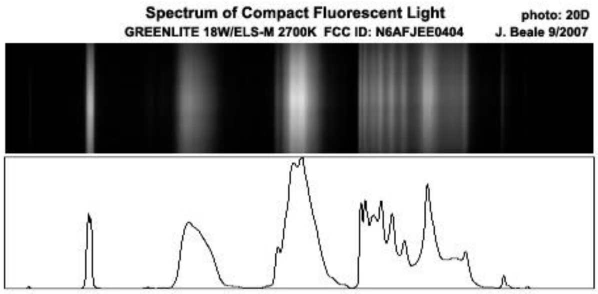


Figure. DIY spectrometer that could serve to keep visitors engaged beyond the visit (left) and spectrum of a 2700 K light (right).

A.4.6 Marker Exhibition

To have better measurements, instead of a single exhibit, ideally we could have a whole exhibition comprising several exhibits. To minimize the need of space and make it a portable exhibition, all exhibits would fit on a table where a dark environment would be artificially created, as in the following image.

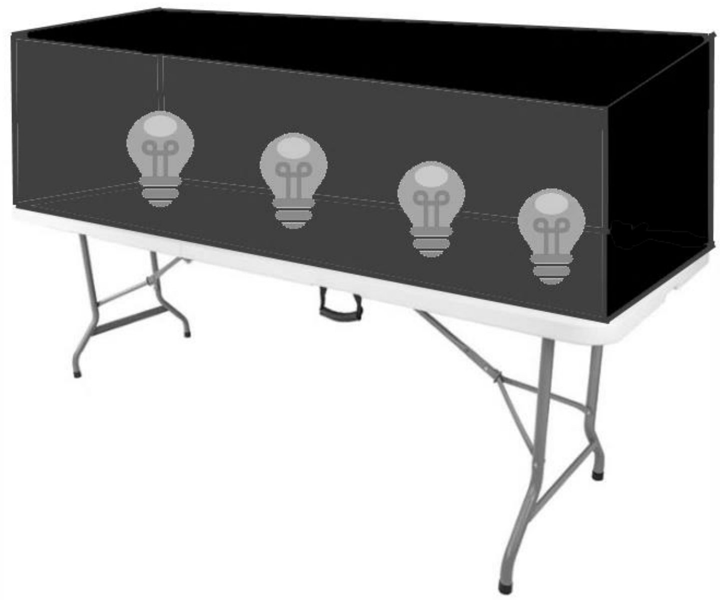


Figure. Portable array of exhibits.

Exhibit A would use diffraction grating glasses with linear divisions of 1000 lines per mm. These glasses would allow to see the spectrum of several light sources: a light bulb, a

CFL lamp, a white LED and an RGB LED. The idea was to show the ‘white’ is actually a rainbow.

White LED Compact Fluorescent Light (CFL) Lamp

Figure. Diffraction glasses (up-left), the sun seen through the glasses (up-right), a white LED with its spectrum (down-left) and a white CFL lamp with its spectrum (down-right).

Exhibit B would show that different colours can be produced from different sets of primary colours and explain why RYB was chosen as primary colours in painting, RGB for electronic screens and CMY(K) for printing.

A desktop version of *Coloured Shadows* would be built, but instead of having three sources of red, green and blue, the three sources would be white. With a set of photography filters, it could be chosen what three colours to use and see what combinations can be created with each set. This exhibit would also allow to explain how the human eye works.

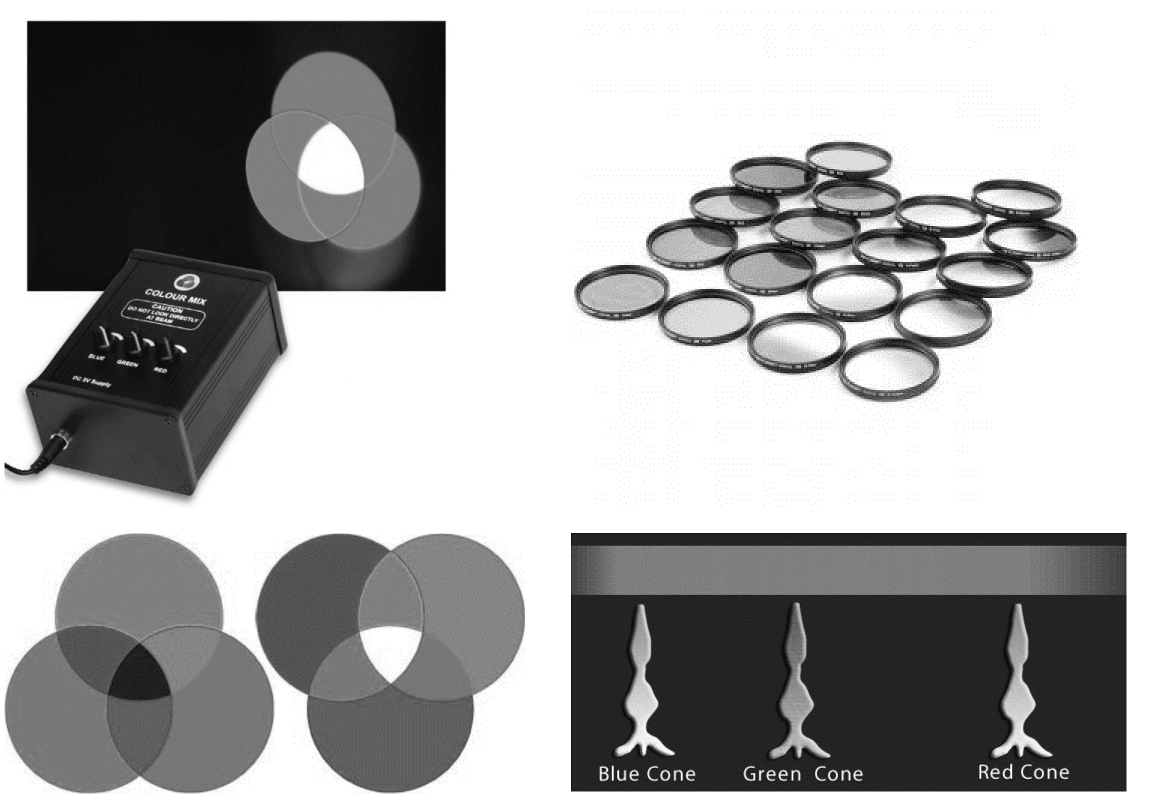


Figure. Desktop version of *Coloured Shadows* (up-left), photography filters (up-right), sets of primary colours (down-left) and human eye's cones (down-right).

Exhibit C would trick the eye in colour production by producing an electronic Newton's Disc. The following image shows the design. LEDs of different colours would be placed on a CD that rotates at user-defined speed, creating the illusion of different colours. A set of switches would allow the user to define what colours to turn on and off, creating their own combinations.

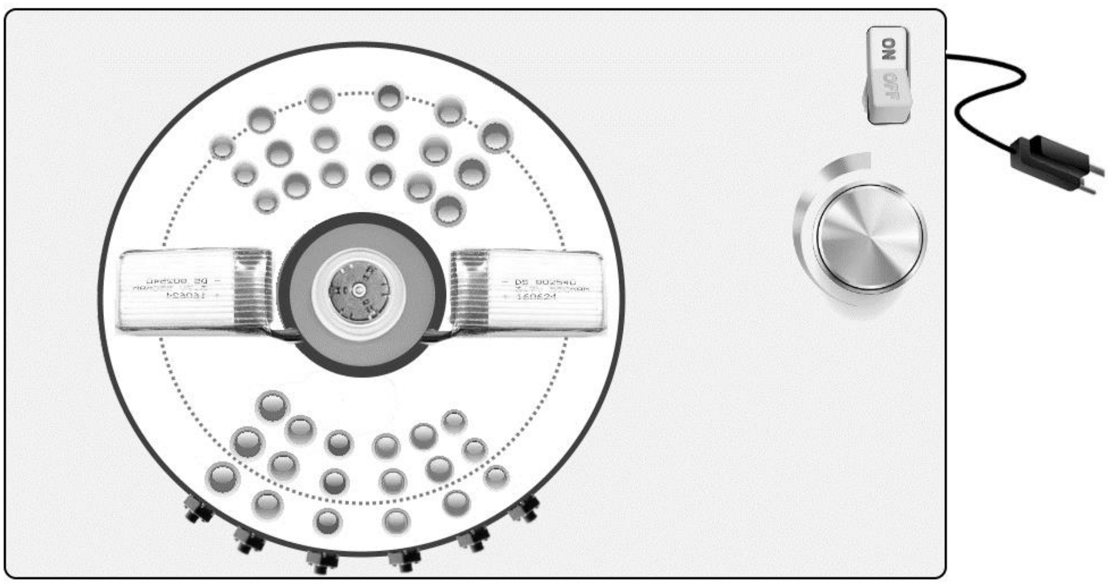


Figure. Electronic version of Newton’s Disc designed by the author.

Exhibit D would use a handheld microscope to show visitors how any image on smartphones and tablets and screens are just a composition of red, green and blue pixels.

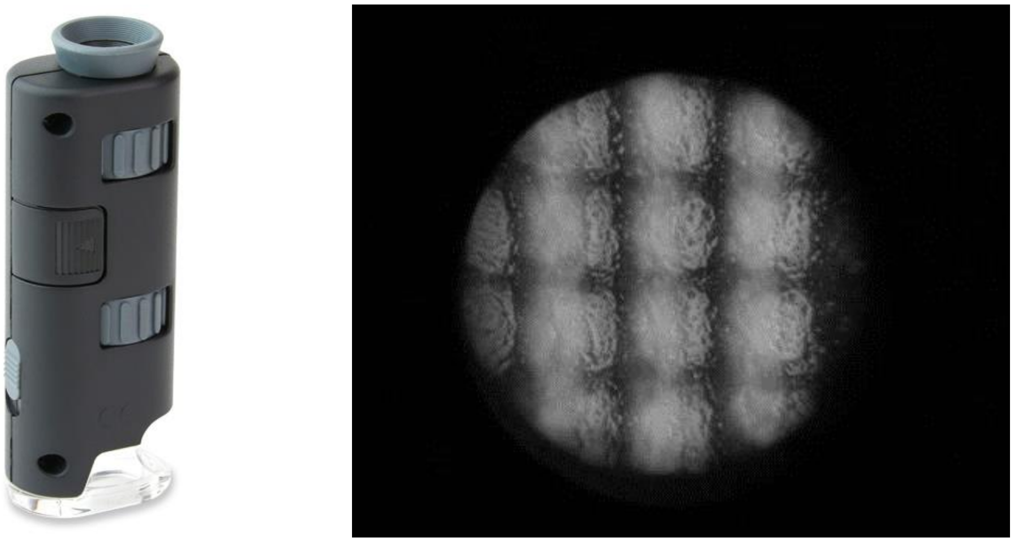


Figure. Handheld microscope (left) used to take a picture of the screen of the author’s smartphone increased 100x.

A.4.7 Bean Poll

A bean poll is an easy way to collect large amounts of data. To make more enticing, visitors would place balloon lights in funnels, depending on the option they want to vote for. The light would roll down through a clear twisted tube that ends in a collector. The colour of the light would be a code that determined gender and age group.

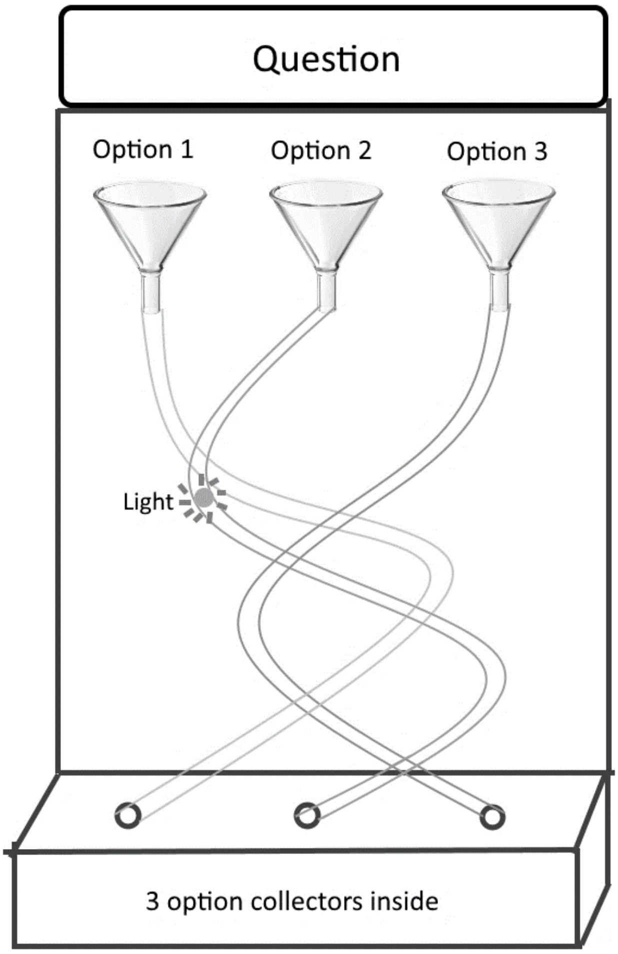


Figure. Balloon lights (left) to place in the Bean Poll (right).

A.4.8 Word Search Puzzle

The idea was to take an individual well-known game of Word Search, but the puzzle contained three 10-word sets (30 words in total). One set was formed of common words that did not appear on any of the Light Zone's panels. One set of common words that appeared on the Light Zone's panels. One set of uncommon, science-related words that appeared on the Light Zone's panels. All three sets were created such that need to have not only the same number of words, but the same total number of characters, trying to keep all of the words with a similar number. After the pilot, the chosen words were as follows.

Common Words not on Panels: *penguin, brother, magazine, movie, ocean, challenge, vacation, character, morning, winter.*

Common Words on Panels: *shadow, universe, rainbow, sculpture, camera, vitamin, liquid, aurora, colour, experiment.*

Uncommon Words on Panels: *spectrum, frequency, sodium, radiation, prism, infrared, plasma, electron, magnet, newton.*

The total number of characters in each set is 71. The minimum/maximum of characters of words in each set is 6/9, 6/10 and 5/9 respectively.

With those 30 words, a 20x20 Word Search Puzzles were created automatically using http://tools.atozteacherstuff.com/word-search-maker/

This method was conducted before and after, under the premise that uncommon science words would be easier to detect after the sleepover. An important instruction was to consider only words of at least five letters (to avoid detection of unwanted words).

The method was popular, but couldn't detect a significant change. It is likely that the problem was the hyperactivity of children during the sleep over (see Chapter 3).

A new index was created by the author and defined as

$$Focus\ Rate\ (FR) = \frac{\%\ Election}{\%\ Probability}$$

where *% Election* is the percentage of words from a group *i* of participants that were found among the total number of found words $\%\ Election, i = \frac{n_{group\ found,i}}{\sum n_{group\ found,i}}$. *% Probability* is the percentage of words from a particular group that would be expected to be randomly found from the pool of available words $\%\ Probability, i = \frac{n_{available\ in\ the\ group,i}}{\sum n_{available\ in\ the\ group,i}}$.

In the pre, all groups have 10 words available (30 in total). Words that were found in the pre are not considered "available" in the post. Therefore, the formula for FR can be rearranged as

$$FR_{i,pre} = \frac{3n_{i,pre}}{\sum n_{i,pre}}$$

$$FR_{i,post} = \frac{(30 - \sum n_{i,pre})n_{i,post}}{(10 - n_{i,pre})\sum n_{i,post}}$$

where *i = [1,2,3], ni,pre* = words found from group i during the pre, *ni,post* = words found from group I during the post.

The following image is an example of a Word Search puzle.

E	W	K	T	B	R	O	T	H	E	R	A	H	T	V	M	S	I	R	P
T	J	F	J	R	W	V	X	I	L	L	I	Q	U	I	D	A	C	G	F
X	V	B	F	E	U	B	E	U	W	G	I	L	M	V	O	Y	O	H	S
Q	A	L	F	T	Y	E	I	D	R	N	H	X	B	N	M	L	L	C	D
F	C	T	G	N	F	U	V	N	O	I	T	A	I	D	A	R	O	A	C
R	A	G	N	I	N	R	O	M	X	U	K	L	O	M	G	K	U	H	C
E	T	B	T	W	I	A	M	N	T	V	Y	T	H	F	A	O	R	V	H
Q	I	E	E	F	U	P	N	I	U	G	N	E	P	U	Z	K	N	G	A
U	O	X	N	X	J	I	O	N	N	Q	X	J	D	I	I	N	P	O	R
E	N	P	G	T	K	Z	T	E	I	R	R	Y	Q	N	N	O	Q	V	A
N	V	E	A	O	A	N	E	W	V	O	R	P	I	F	E	A	W	I	C
C	P	R	M	N	L	A	G	T	E	R	W	W	A	R	R	R	X	T	T
Y	P	I	S	B	W	E	N	O	R	A	I	O	P	A	Z	E	S	A	E
W	L	M	S	A	O	C	E	N	S	P	E	C	T	R	U	M	Q	M	R
L	A	E	I	K	B	O	L	G	E	D	C	V	I	E	E	A	Q	I	H
B	S	N	M	O	N	P	L	S	O	D	I	U	M	D	N	C	O	N	R
E	M	T	T	G	I	N	A	R	W	F	N	O	R	T	C	E	L	E	Y
X	A	M	C	G	A	P	H	W	O	D	A	H	S	C	M	Y	W	S	A
C	U	U	D	I	R	R	C	P	S	U	R	F	D	K	K	T	Y	B	I
U	R	L	S	E	R	U	T	P	L	U	C	S	A	U	R	O	R	A	N

Figure. Word Search puzzle of 20x20.

A.4.9 Open Questions

Some other pilots were conducted with open questions. In one, children received diffraction glasses to look at different sources of light and they were asked, before and after the sleepover, to explain why they look different. In another pilot, children were asked ‘What different kinds of light can you think of?’ and ‘What makes them different’. In yet another pilot, visitors were asked to find the colours that can be combined to make up white and justify their answer (see the following image).

0. Wristband number: _________

1. Circle the colours you would combine to produce White.
(if you can think of more than one set of colours, use a different felt pen for each one)

Brown	Green	Yellow	Purple
Orange	Indigo	Cyan	Pink
Red	Violet	Gray	Blue

2. Please, tell us about why you chose those colours.

Figure. Instrument to detect if visitors are more likely to know that reg, green and blue are primary colours after the sleepover.

None of the instruments were able to detect significant changes in children. The last instrument, shown above in the image, is of special interest. In surveys it was shown that visitors significantly increase their knowledge about what colours are primary colours (see Chapter 4). The fact that the instrument couldn’t detect any changes, backs the idea that the problem was not the instruments, but the unbounded conditions of sleepovers.

Appendix B. Ethics Approvals

B.1 Human Ethics Committee of the University of Otago. Approval 17/062

17/062

Academic Services
Manager, Academic Committees, Mr Gary Witte

25 May 2017

Professor N Longnecker
Centre for Science Communication
133 Union St East

Dear Professor Longnecker,

I am again writing to you concerning your proposal entitled "**Effectiveness Redevelopment of Discovery World at the Otago Museum**", Ethics Committee reference number **17/062**.

Thank you for your email of 23rd May 2017 with attached response and revised documentation addressing the issues raised by the Committee.

On the basis of this response, I am pleased to confirm that the proposal now has full ethical approval to proceed.

Approval is for up to three years from the date of this letter. If this project has not been completed within three years from the date of this letter, re-approval must be requested. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing.

The Human Ethics Committee asks for a Final Report to be provided upon completion of the study. The Final Report template can be found on the Human Ethics Web Page

http://www.otago.ac.nz/council/committees/committees/HumanEthicsCommittees.html

Yours sincerely,

Mr Gary Witte
Manager, Academic Committees
Tel: 479 8256
Email: gary.witte@otago.ac.nz

c.c. Professor L S Davis Director Centre for Science Communication

B.2 Human Ethics Committee of the University of Otago. Approval 17/062 (amendment)

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

17/062

Academic Services
Manager, Academic Committees, Mr Gary Witte

17 April 2018

Professor N Longnecker
Centre for Science Communication
133 Union St East

Dear Professor Longnecker,

I am again writing to you concerning your proposal entitled "**Effectiveness Redevelopment of Discovery World at the Otago Museum**", Ethics Committee reference number **17/062**.

Thank you for your email of 16th April 2017 with request for amendment attached.

The Committee notes that you wish to make the following changes to the study:

1. update the name of the centre on all material from Discovery World to Tūhura,
2. give participants a small token of appreciation, such as a glow-in-the-dark star,
3. take pictures of some participants at Tūhura (with parental consent)
4. note verbatim comments of visitors who chat with the researcher on site. You note that some interesting comments were made to the researcher while delivering the Discovery World surveys but that these comments were not recorded as you had neither approval nor consent to do so. You note that you will now have Information Sheets and Consent Forms ready for these cases.

The Committee further notes the modifications to the pre- and post-survey, focus group questions, replacement of the Mind Map with a Self-Questioning, Self-Answering activity and the addition of the use of the Questions and Answers board and Bean Poll.

The Committee accepts and approves the amendment.

While approving the amendment, please provide the Committee with a copy of the parental consent form allowing photographs to be taken of participating children.

Your proposal continues to be fully approved by the Human Ethics Committee. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing. I hope all goes well for you with your upcoming research.

B.3 Human Ethics Committee of the University of Otago. Approval D17/186

UNIVERSITY OF OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

D17/186

Academic Services
Manager, Academic Committees, Mr Gary Witte

9 June 2017

Professor N Longnecker
Centre for Science Communication
133 Union St East

Dear Professor Longnecker,

I am writing to confirm for you the status of your proposal entitled "**Staff Views on the Discovery World Redevelopment**", which was originally received on May 31, 2017. The Human Ethics Committee's reference number for this proposal is **D17/186**.

The above application was Category B and had therefore been considered within the Department or School. The outcome was subsequently reviewed by the University of Otago Human Ethics Committee. The outcome of that consideration was that the proposal was approved.

Approval is for up to three years from the date of HOD approval. If this project has not been completed within three years of this date, re-approval must be requested. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing.

Yours sincerely,

Gary Witte

Mr Gary Witte
Manager, Academic Committees
Tel: 479 8256
Email: gary.witte@otago.ac.nz

B.4 Māori Research Advisor of the University of Otago. Approval 5697_19577

NGĀI TAHU RESEARCH CONSULTATION COMMITTEE

TE KOMITI RAKAHAU KI KĀI TAHU

Tuesday, 02 May 2017.

Professor Nancy Longnecker,
Centre for Science Communication,
DUNEDIN.

Tēnā koe Professor Nancy Longnecker,

Effectiveness Redevelopment of Discovery World at the Otago Museum

The Ngāi Tahu Research Consultation Committee (the committee) met on Tuesday, 02 May 2017 to discuss your research proposition.

By way of introduction, this response from The Committee is provided as part of the Memorandum of Understanding between Te Rūnanga o Ngāi Tahu and the University. In the statement of principles of the memorandum it states "Ngāi Tahu acknowledges that the consultation process outline in this policy provides no power of veto by Ngāi Tahu to research undertaken at the University of Otago". As such, this response is not "approval" or "mandate" for the research, rather it is a mandated response from a Ngāi Tahu appointed committee. This process is part of a number of requirements for researchers to undertake and does not cover other issues relating to ethics, including methodology they are separate requirements with other committees, for example the Human Ethics Committee, etc.

Within the context of the Policy for Research Consultation with Māori, the Committee base consultation on that defined by Justice McGechan:

"Consultation does not mean negotiation or agreement. It means: setting out a proposal not fully decided upon; adequately informing a party about relevant information upon which the proposal is based; listening to what the others have to say with an open mind (in that there is room to be persuaded against the proposal); undertaking that task in a genuine and not cosmetic manner. Reaching a decision that may or may not alter the original proposal."

The Committee considers the research to be of interest and importance.

As this study involves human participants, the Committee strongly encourage that ethnicity data be collected as part of the research project as a right to express their self-identity. That is the questions on self-identified ethnicity and descent, these questions are contained in the latest census.

The Committee asks if the researchers have consulted with the Māori Advisory Group of the Otago Museum or had contact with Rachel Wesley, Otago Museum's Curator – Māori.

The Committee suggests dissemination of the research findings to relevant National Māori Education organizations, Toitu te Iwi at Te Rūnanga o Ngāi Tahu and to the Office of Māori Development, University of Otago regarding this study.

NGĀI TAHU RESEARCH CONSULTATION COMMITTEE
TE KOMITI RAKAHAU KI KĀI TAHU

We wish you every success in your research and the Committee also requests a copy of the research findings.

This letter of suggestion, recommendation and advice is current for an 18 month period from Tuesday, 02 May 2017 to 2 November 2018.

Nāhaku noa, nā

PP NTRCC

Mark Brunton
Kaiwhakahaere Rangahau Māori
Research Manager Māori
Research Division
Te Whare Wānanga o Otāgo
Ph: +64 3 479 8738
Email: mark.brunton@otago.ac.nz
Web: www.otago.ac.nz

The Ngāi Tahu Research Consultation Committee has membership from:

Te Rūnanga o Ōtākou Incorporated
Kāti Huirapa Rūnaka ki Puketeraki
Te Rūnanga o Moeraki

Appendix C. Selected Exhibits

C.1 Selection process

To decide what exhibits use as examples, three sources were considered: focus groups, Staff interviews and surveys (responses to "It was cool to learn about…"[175] and "Do you miss anything from Discovery World?").

The following table shows the list of exhibits considered by staff as their personal favourite and least favourite exhibits. The Tropical Forest and the planetarium are excluded, but the Science Show is kept as exhibit.

Table
Staff's most favourite and least favourite exhibits

Discovery World			
Favourite Exhibits		**Least Favourite Exhibits**	
Kinetic Sculpture	6	Air Hockey Table	6
Plasma Room	5	Foosball Table	5
Frozen Shadows	3	Mind Ball	3
Mind Ball	3	Paint with Light	3
Ray Table	3	Torsion Wave	3
Reaction Timer	2	Icy Bodies	2
Other Exhibits	13	Piano	2
		Other Exhibits	7
Tūhura			
Favourite Exhibits		**Least Favourite Exhibits**	
The Void	6	Dancing with Lights	3
Chicken Embryo	5	Mushrooms	3
Torque Table	5	Oxygen Bubbles	3
Flight Zone	4	Slide	3
Dams Table	3	Blue and Red Buttons	2
Magnetic Sand	3	Dams Table	2
Microeye	3	Newton's Prism	2
Gravity Wall	2	Shape Builder	2
Plasma Tube	2	Specimens Drawers	2
Slide	2	Topography Table	2
Sound Bite	2	Other Exhibits	8
Other Exhibits	10		

The following figure shows the most mentioned exhibits in Discovery World regarding "It was cool to learn about…". Only those that obtained at least four mentions are displayed individually. *Plasma Room* was, technically speaking, not a single exhibit, but a dark room with four exhibits inside (*Jacobs Ladder, Neon Lightning, Plasma Plates* and *Plasma Tube*), but generally visitors and staff identified it as a single one and therefore it was taken so in this research.

[175] 'Have your say! Let us know what you liked or didn't like of Tūhura's exhibits or how they can be improved' and 'Can you give an example of something you learnt?' were included in Tūhura's code by topic.

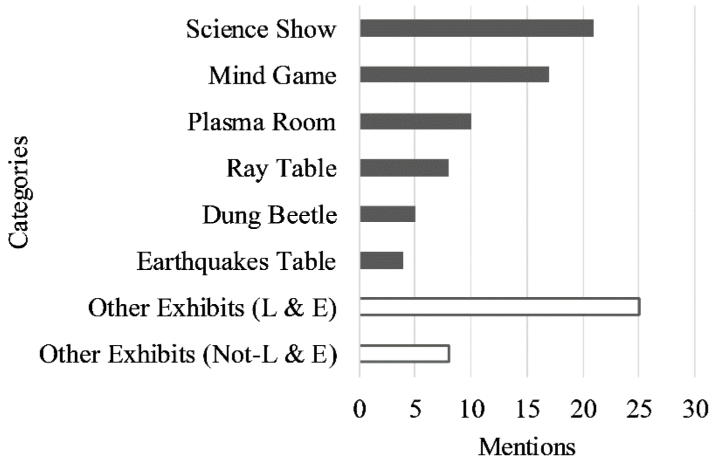


Figure. Most mentioned exhibits in responses to "It was cool learning about..." in Discovery World (173 respondents, 214 category mentions). L & E stands for Light and Electromagnetism exhibits. White bars do not correspond to a single specific exhibit.

The following figure shows the most mentioned exhibits Tūhura visitors say they miss from Discovery World. The sample only includes visitors that visited Discovery World before it closed. The cut-off is 11 mentions.

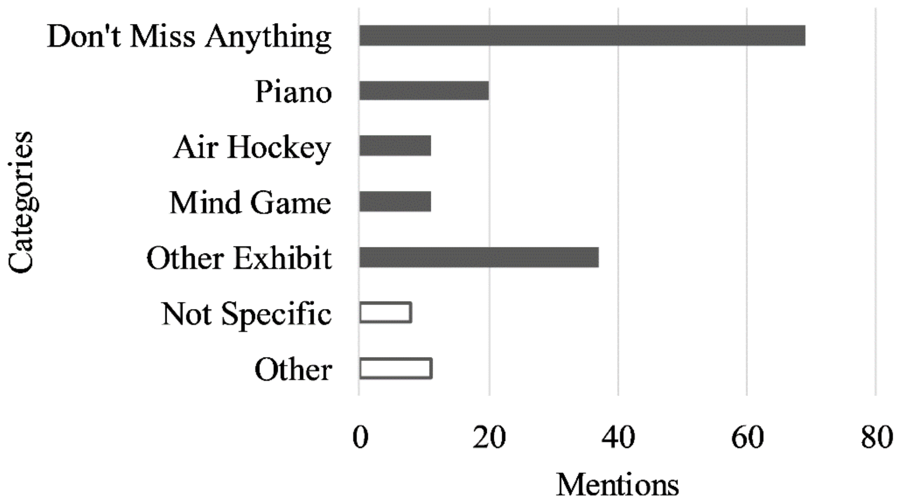


Figure. Exhibits mentioned in 'Do you miss anything from Discovery World?' (220 respondents, 230 category mentions). White bars do not correspond to a single specific exhibit.

The following figure shows the most mentioned exhibits for Tūhura regarding "It was cool to learn about…". The cut-off of 7 mentions. Exhibits with 4 to 6 mentions were: *Team Pac-Man*, *Magma Chamber*, *Microeye*, *Touch-Sensitive Plant*, *Order of Arthropods*, *Gravity* *wall + Vacuum drop*[176], *Velvet Hands, Predator vs Prey + Human Skulls + Skeleton Station*[177]. Exhibits with 1 to 3 mentions were: *Floating in Copper, Ultraviolet Camera, Funhouse Mirrors, Seismometer, Communication conundrums, Brine Shrimp Ballet, Startle Reflex, Platypus, Red Team vs Blue Team, Rocks and Dancing with Lights*. Some exhibits didn't get any mentions: *Light Island, Bacteria Walls, Oxygen Bubbles, Mushroom Morphology and Face Blender*.

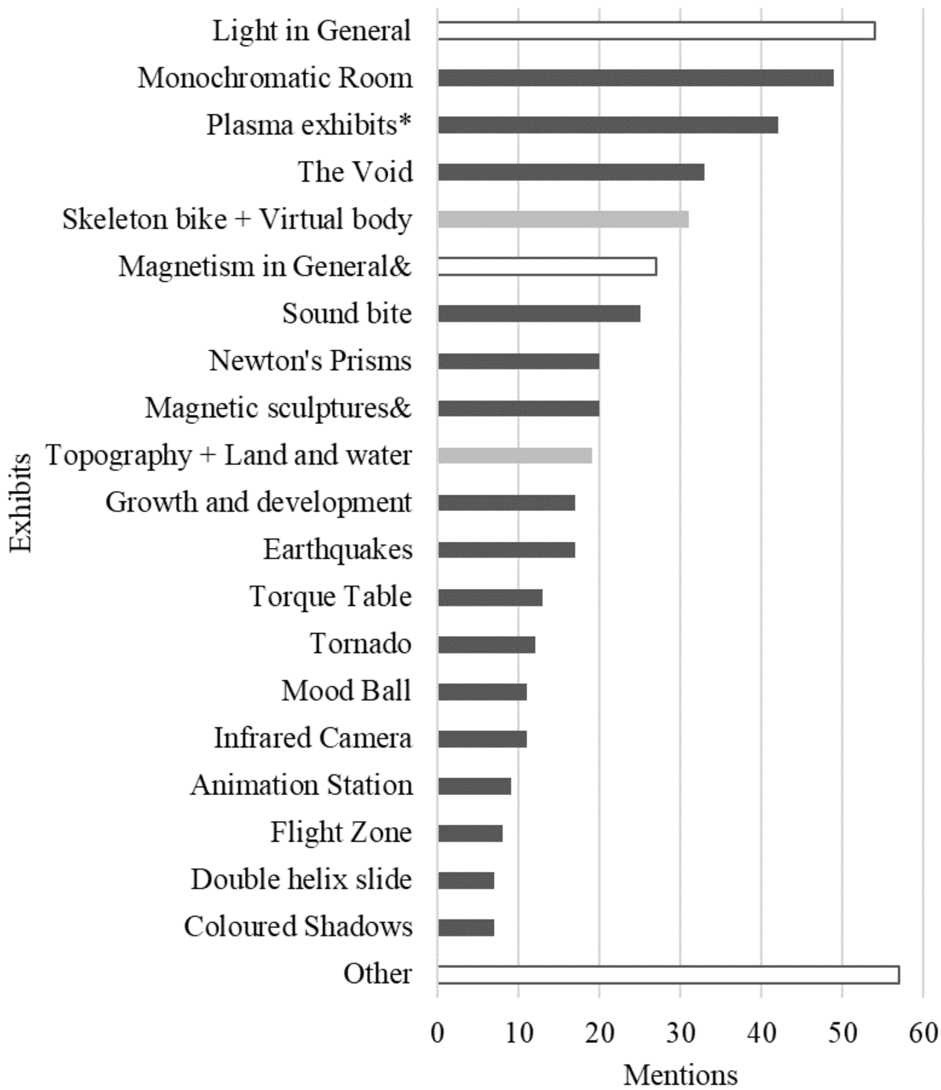


Figure. Most mentioned exhibits in responses to "It was cool learning about..." in Tūhura (610 respondents, 778 category mentions). White bars do not correspond to a single specific exhibit. Light grey bars include two exhibits in one. (*) is technically two exhibits, but so similar, that they are considered as a single exhibit. Magnetism in General (&) most likely refers to *Magnetic Sculptures* (&), in which case, this exhibit would lie just behind Monochromatic Room.

[176] Exhibits marked with * means that it is more than one exhibit, but it is too difficult to distinguish when visitors talk about one and when about the other.
[177] Same as the previous footnote.

C.2 Selected exhibits

Based on the previous section, the following table presents a list of exhibits holistically selected. According to the best judgment of the author, they are representative of three levels of science communication. Although exhibits from several areas were included, preference was given to exhibits related to light and electromagnetism. 'Need a different approach' are exhibits that seem not to be contributing to learning science in their current form. It does not mean necessarily that the exhibit itself is not appropriate, but the way it is presented needs to be clearer. 'On track' are those that are contributing but can become better. 'Gold standard' are those that can be considered the best examples of science communication[178].

Table

Selected exhibits in three levels of their levels of science communication

Need a different approach	**On track**	**Gold standard**
Air Hockey Table	Ray Table	Plasma Room
UV Camera	IR Camera	Mind Ball
Floating in Copper	The Void	Torque Table
Light Island	Coloured Shadows	Magnetic Sculptures
Blue Team vs Red Team	Newton's Prism	Monochromatic Room
DNA Slide	Dancing with Lights	Frozen Shadows
Shark Hologram	Chicken Embryo	Sound Bite
	Vacuum Drop	Kinetic Sculpture

[178] It doesn't mean they are perfect, everything is perfectible.

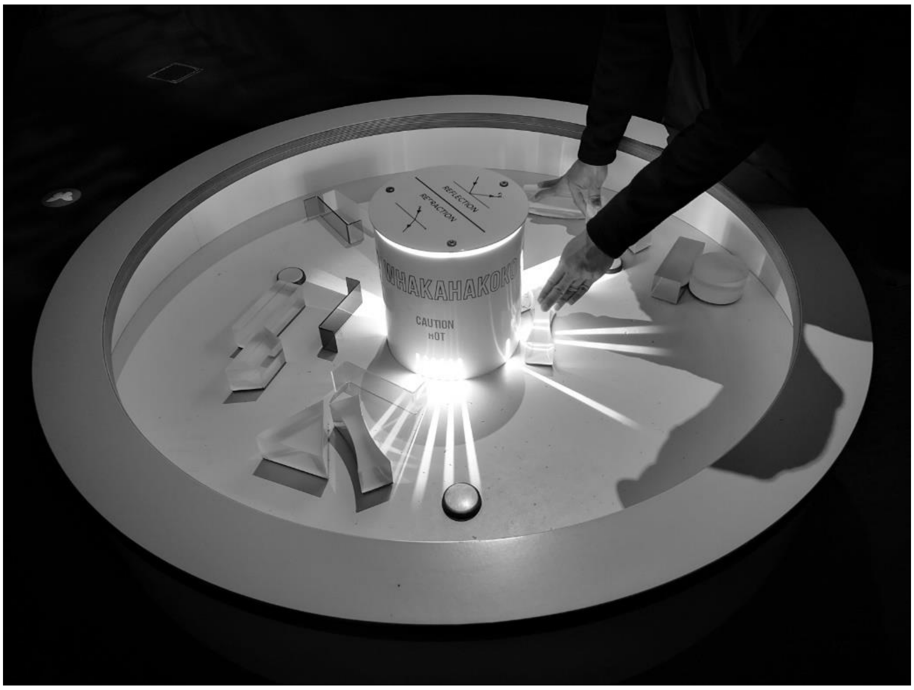
C.3 Exhibits that need a different approach

Air Hockey Table. A commercial air hockey table.

UV Camera. An ultraviolet camera displays what it detects in grayscale on a big screen (right in the following picture).

Floating in Copper. In-between a sandwich of two copper section lies a metal ball. The user is supposed to make the ball levitate using an external magnet.

Light Island. White, red, green and blue lights are emitted from a central source in form of collimated beams. Clear shapes are available to test how they affect light.



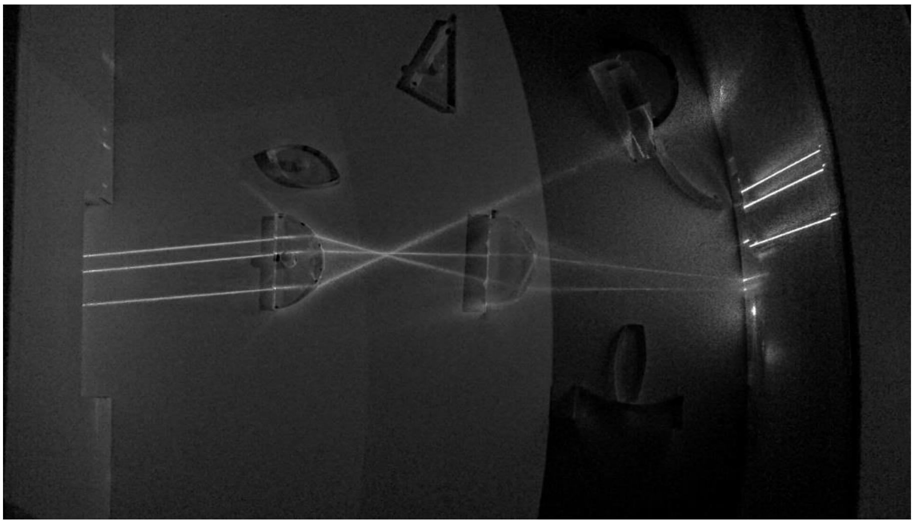
Blue Team vs Red Team. Visitors press either a blue or a red button as fast as they can. A screen displays how many times each colour has been pressed.

DNA Slide. A big slide resembles the shape of the DNA’s double-helix.

Shark Hologram. From the right spot, a frame displays a shark hologram. Not visible from other positions.

C.4 Exhibits that are on track

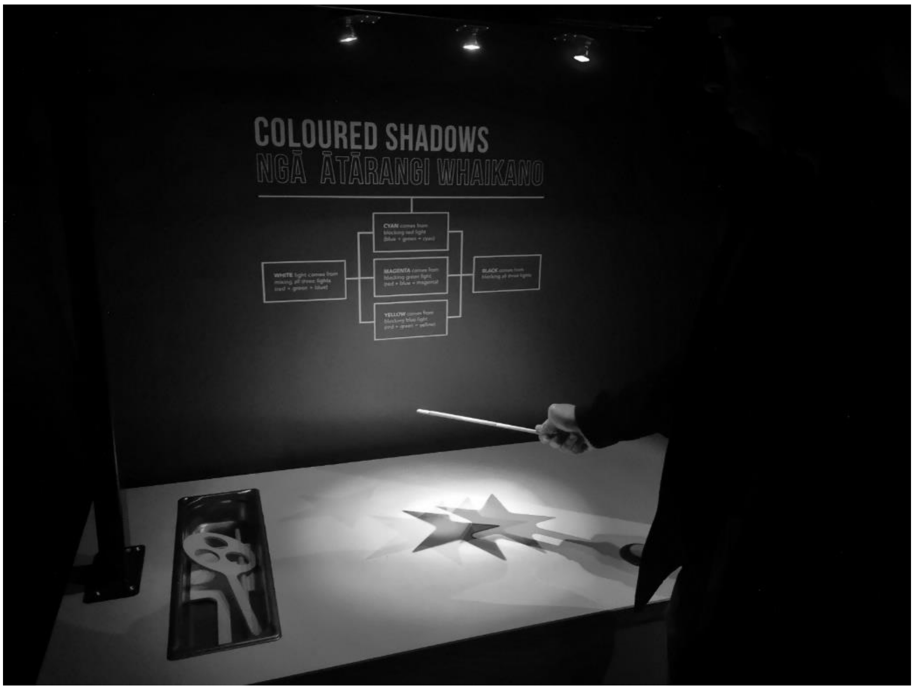
Ray Table. Three red lasers shone from a hidden source. The beams were parallel, but with the help of a series of clear shapes it was possible to diffract.



IR Camera. An infrared camera displays what it captures on a big screen (left). A colour code on the top of the screen shows what temperature corresponds to what colour.

The Void. It is a room where colourful lights turn on and off rhythmically while relaxing sounds. Mirrors on the walls make the lights repeat themselves infinitely.

Coloured Shadows. Three knobs allow to grade the intensity of red, green and blue lights independently. Several shapes are available to produce coloured shadows.



Newton’s Prism. A source of white light passes through a rotating prism. The user turns the prism to make the beam pass through a second rotating prism and finally be projected onto a white screen where the rainbow is displayed.

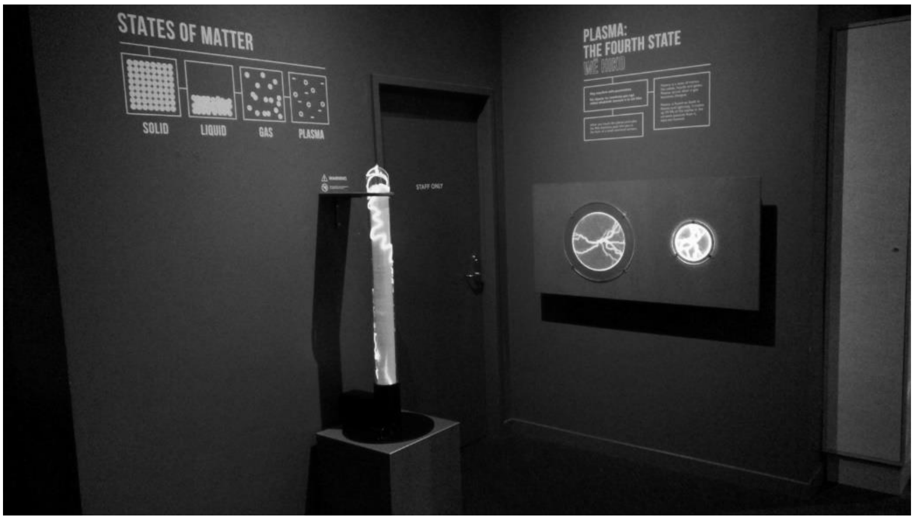
Dancing with Lights. A Kinect sensor detects the user and then projects the image onto a wall. The image is modified to be presented in psychedelic colours and it is updated every certain time, with the previous images slowly fading away.

Chicken Embryo. Several stages of live chicken embryos are presented. Each one has a magnifying glass to appreciate the details. The hearth can be clearly seen beating, especially in the last stage. The embryos need to be replaced daily.

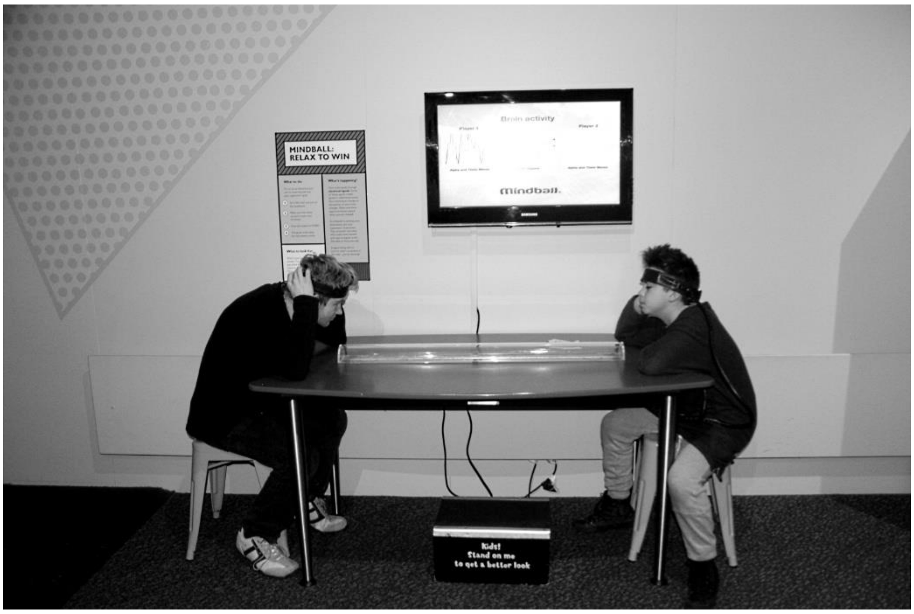
Vacuum Drop. Inside a clear tube there is a feather and a small ball. Air in the tube can be extracted to compare the speed at which these objects fall with and without air.

C.5 Exhibits that are gold standard

Plasma Room / Plasma: the fourth state and States of Matter. *Plasma Room* was technically not a exhibit, but a dark room with four exhibits: *Jacobs Ladder, Neon Lightning, Plasma Plates* and *Plasma Tube*. The latter two were kept at Tūhura under the names of Plasma: the fourth state and States of Matter, respectively. The following images are of the sign at the entrance of the Plasma Room and of the two exhibits that were kept.

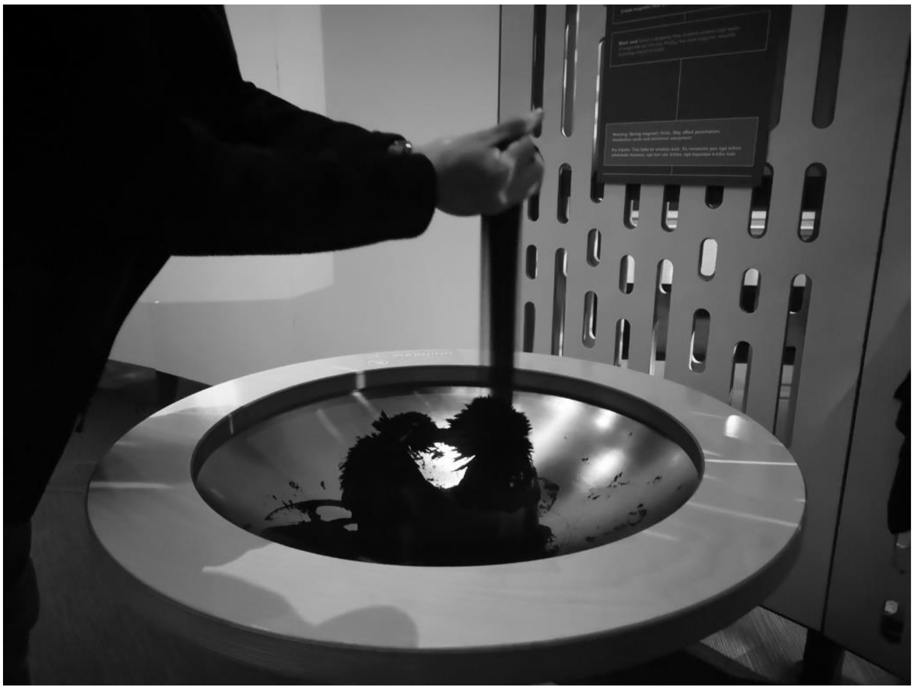


Mind Ball. Two participants place headbands to detect electrical signals in the fashion of an electroencephalogram. A computerized system determined what participant was more relaxed and moves a little ball inside a clear tube between the participants. Eventually, the most relaxed participant pushes the ball all the way to the opponent's side and wins.



Torque Table. A large metal discs rotates when a visitor approaches. There is a lot of shapes to place onto the disc to discover how they react to circular motion.

Magnetic Sculptures. A bowl contains two strong permanent magnets and magnetic sand.

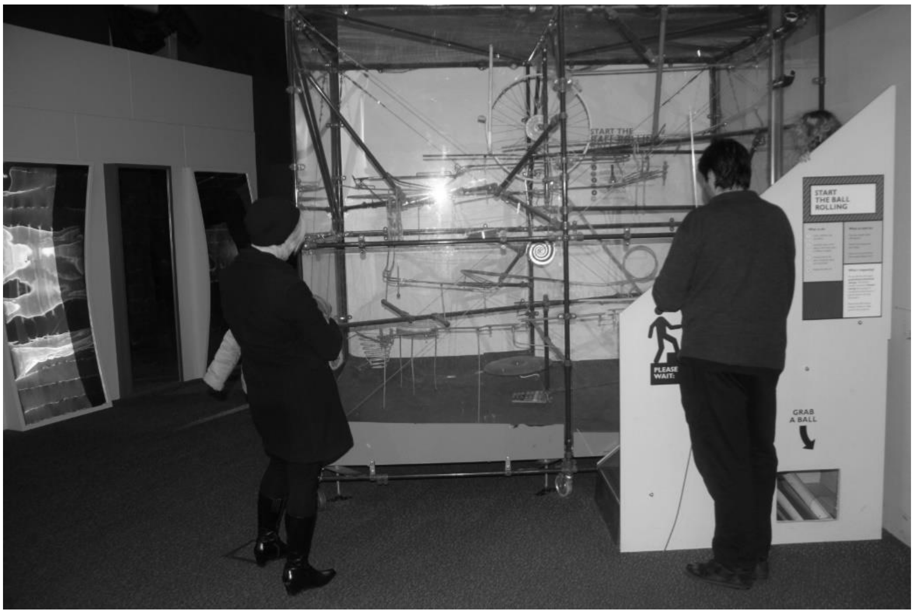


Monochromatic Room. A room filled with colourful items is illuminated with monochromatic light. All items look yellow-ish and dull, but when illuminated with one of the flashlights available, their true colours show up.

Sound Bite. A metal rod is vibrating with music. When the visitor bites it, the sound waves are transmitted to the jaw bones and the visitor can hear the music directly in their head.

Frozen Shadows. Visitors pose in front of a photosensitive wall. A flash then illuminates then room and the shadows of the participants remain on the wall.

Kinetic Sculpture. It is a big Goldberg machine. The user places a ball up and it rolls down by gravity.



C.6 Around the exhibits

The Greatest Science Show in the History of the World, Ever! Better known just as 'the Science Show', it was an interactive show conducted at Discovery World. The Science Show was highly regarded by staff. "My favourite thing in Discovery World at the moment isn't an exhibit, it's the Science Shows" (DM3), because

> I think the interactive nature of the show is very inspirational. I think we've got a talented team of science communicators who run the shows … So, from my perspective, it's the personality of the shows that engage the audience with the science and I think that's a very successful way of doing things. (DM3)

This type of activities are important because children associate ideas of science more readily with shows than with exhibits (Dicks, 2013), as confirmed by the following visitor comment. "The science show was spot on for the kids. Excellent!" (DW, M, 19-40). But not only children, adults also learn science at the Science Show. "[I learnt] how you can hold flames in your hand without them burning if they are hydrogen gas caught in detergent bubbles" (DW, F, 41+). Unfortunately,

> one element of Discovery World that we deliberately cut out was the Science Shows. In the old version of Discovery World, the exhibits were so pathetic that we had to kind of do a lot to kind of show value, or to add value to the experience. We spent a lot of time developing science shows inside the space. And the nature of Tūhura means that we can't really do those shows inside the space anymore, so the formal twenty-minute presentation of the science show, we don't do in Tūhura anymore, we do it outside of that space. (DM3)

Some Science Communicators miss having the Science Shows in Tūhura. "Merely having a dedicated theatre is what I miss" (SC7).

> I'm still disappointed that there is no space in where we can do regular science shows, because it's definitely something that we can provide … It's not a matter of giving people what they ask for, I think it's a matter of giving people what we can, and we can do something that we are not giving people now, and that's a shame. (SC5)

Anecdotally, a man in his 70s approached the author to chat, ignoring he was a researcher. He's from Sydney, Australia, and all the years, he, his wife and his grandson come to Dunedin. Every single year they come to the science centre. He liked Tūhura, but he was very disappointed by the absence of the Science Show. "Bring back the liquid nitrogen and hydrogen explosions!!!!!" was the note he left at Tūhura's Q&A board.

Fortunately, the museum is already preparing the return of the Science Show to Tūhura.

Figure. Discovery World’s Science Show in action.

Beautiful Science Gallery. This space is an interactive gallery (the images projected on the walls can be moved with the hand thanks to a laser system that records the hands movements). It is placed between Tūhura (behind the left wall in the image) and the planetarium (behind the right wall in the image).

Figure. Beautiful Science gallery.

Panels. Labels at Discovery World followed a specific pattern that can be seen in the first of the following images. At Tūhura, there is no pattern labels wise. The next images are examples of the large variety of panels that can be found at Tūhura.

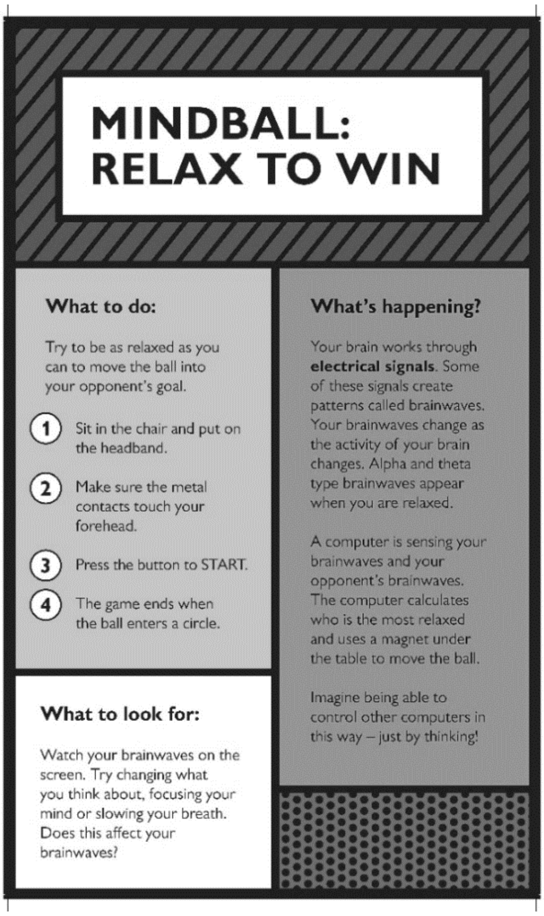


Figure. Typical label at Discovery World.

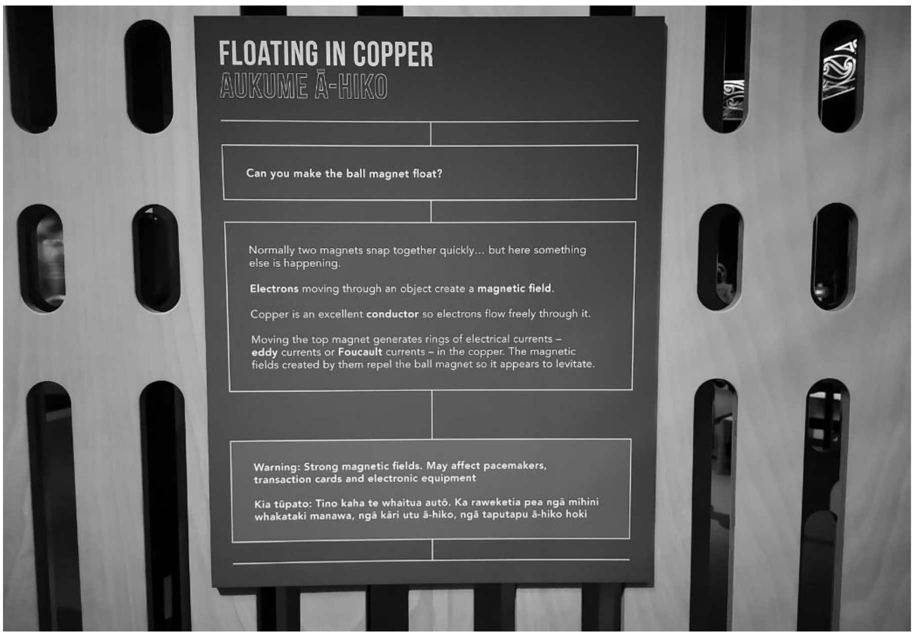


Figure. Example of label (Tūhura).

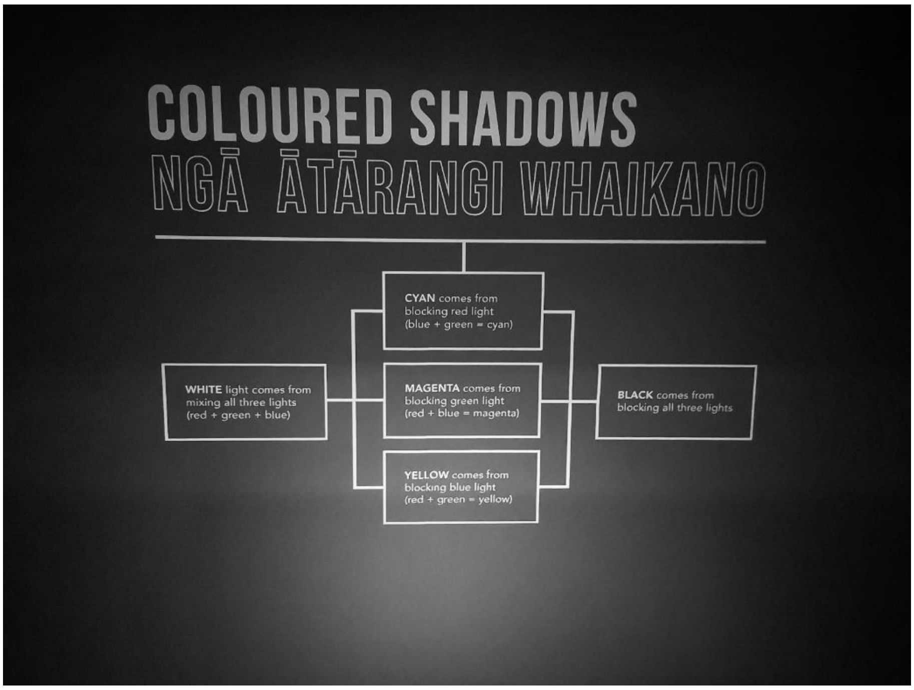


Figure. Example of label (Tūhura).

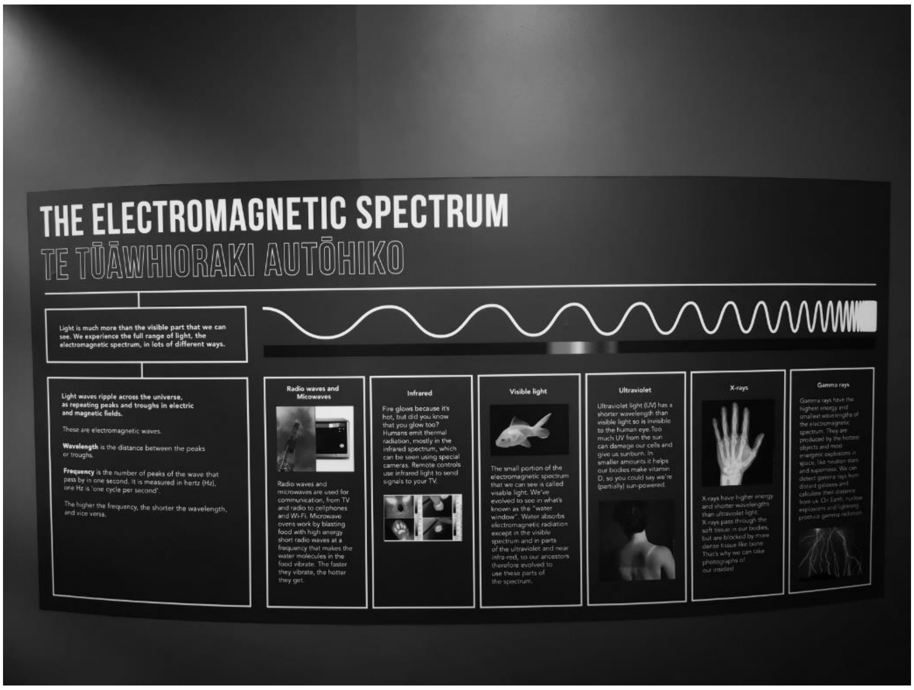


Figure. Label without an associated exhibition (Tūhura).

Figure. Label on the floor (Tūhura).

Appendix D. Data Pre-processing

For ease of referencing, in this appendix, the 'Survey focused on knowledge in Discovery World' will be referred as Survey 1. The 'Survey focused on knowledge in Tūhura' as Survey 2. The 'Survey focused on gender differences in Tūhura' as Survey 3. And the 'Survey focused on self-beliefs in Tūhura' as Survey 4[179].

D.1 Removal of survey drop-outs

Some survey respondents withdrew either explicitly or by not coming back to the post-survey. The number of drop-outs for Survey 1 were 35 of 262 (13%). For Survey 2 were 45 of 535 (8%). For Survey 3 were 24 of 423 (6%). For Survey 4 were 18 of 279 (6%). The number of responses per survey after removing drop-outs was Survey 1: 227. Survey 2: 490. Survey 3: 400. Survey 4: 261. Table 1 shows the demographics of drop-outs (Table).

Table
Demographics of drop-outs. Figures represent absolute numbers of visitors in each category

		Survey 1	Survey 2	Survey 3	Survey 4
Gender	Male	15	16	4	8
	Female	18	27	17	8
	Other/missing	2	2	2	2
Age group	8-12	11	16	2	3
	13-18	7	6	5	4
	19-40	8	8	10	1
	41+	9	13	4	7
	Other/missing	0	2	2	3

D.2 Removal of survey invalid times

Due to a problem with system memory in the iPads, the time required to fill out the survey sometimes overflew. To detect and remove them, a cut-off time was calculated as Q3+5*IQR, where Q3 is the third quartile and IQR is the interquartile range. Negative values were also overflown values that were removed.

Some other times, values were not recorded for different reasons, mainly technical. Some values were also removed manually for different reasons (e.g. someone stopping to do something else for a long time and coming back to finish after a while).

[179] The surveys can be found in Section A.1.

Survey 1 didn’t have overflowing problem, but anyhow the combined pre + post cut-off time was calculated as 1494 seconds (25 min). Only one datum was out of this limit (2069 secs). It was not deleted as the case appears in the notes of the researcher. Twentynine values were not recorded in Survey 1 (15 because the surveys were filled out on paper, six because the pilot version was not still set to record them, and eight because of pre and post were completed on different iPads). One value from visit time was removed manually. It corresponds to an extreme outlier that the researcher had in the notes as someone that left Discovery World to go to the planetarium[180]. Outliers need to be removed to provide an actual reflection of the phenomena examined (Aguinis, Gottfredson, & Joo, 2013). A total of 424 time points remained valid.

For Survey 2, the pre-survey cut-off was 956 seconds (16 minutes) and 1022 seconds (17 minutes) for the post. Six values were removed with the cut-off criteria. Sixty values were not recorded (54 of them were due to an error in the configuration of the piloted survey). Eight invalid data were removed manually. A total of 1397 time points remained valid.

For Survey 3, the pre-survey cut-off was 1035 seconds (17 min) and 811 seconds (14 min) for the post. Thirteen values were removed for not passing the cut-off values. Seven values were not recorded. 20 were deleted manually. A total of 1160 time points remained valid.

For Survey 4, the pre-survey cut-off were 811 seconds (14 min) and 597 seconds (10 min) for the post. Only one datum didn’t pass the cut-offs. Sixty-six times were not recorded because those surveys were on paper. No values were removed manually (no negative values were found). It seems overflowing stopped happening, probably due to a software update.

After the removal of invalid times it is possible to see that surveys were, as expected from design, short. The following table shows the first quartile, median (second quartile) and third quartile of all surveys.

[180] Remember that in Discovery World the planetarium was outside, therefore its visits were not counted in the science centre visit time. In Tūhura, the planetarium is in the same area. It happened sometimes that parents left Discovery World / Tūhura to have coffee or to put money to the parking meter, inflating their visit time, but that was not usual. This is a reason why the median and not the mean is a better indicator.

Table

Median time employed by visitors to fill out surveys before (pre) and after (post) the visit

	Pre-survey (before the visit)			Post-survey (after the visit)		
	Q1	**Mdn**	**Q3**	**Q1**	**Mdn**	**Q3**
Survey 1 (n=220)	276	350	479			
Survey 2 (n=428)	193.25	244	317.5	139	195	279.5
Survey 3 (n=359)	175	234	299	122	158	230
Survey 4 (n=216)	141	177.5	253.5	86.25	122	172.5

Discovery World's Survey 1 didn't record the time in each pre- post-survey independently, but only one as the sum of both.

D.3 Removal of invalid responses by points criteria

Ideally, all respondents would be committed to the survey. In real life, some of them are not and do not answer with their actual opinions. Their responses need to be detected to avoid bias. Examples of characteristics of these respondents are finishing very quickly, skipping sections and answering in patterns (e.g., a complete scale with the same answer or in diagonal). None of these behaviours alone is irrefutable indicative of a non-committed respondent, but several of them might be.

An Excel file was created to detect automatically warning signals and assign points to them according to the following criteria:

a) Respondents with the lowest time needed to finish the survey got 1.5 points if in the lower 2.5% percentile, 1.0 points if in the lower 5% and 0.5 points if in the lower 10%.
b) Sections with at least 3 questions/spaces to answer (e.g., Likert-type scale, Three words to describe the science centre, etc.) received 1.0 points if completely skipped[181], or 0.5 points if at least 50% of the section was skipped.
c) Sections with at least 4 questions/spaces[182] filled out in 51% or more (3 of 4, 3 of 5, 4 of 6, 4 of 7, 5 of 8) received 0.5 points if all answers were the same.
d) Sections with at least 4 questions/spaces answered in 51% or more (3 of 4, 3 of 5, 4 of 6, 4 of 7, 5 of 8) received 0.5 points if items were answered in a diagonal or arrow (i.e., there was a geometric pattern).

[181] In Discovery World, the survey was programmed to stop respondents from skipping sections; this was corrected in Tūhura (they should be allowed to skip them if they so desire, according to the Information Sheet).
[182] Questions of 3 items were not considered here because it might be too strict to think of a pattern with only 3 items.

e) Open questions (except Three words to describe the science centre), such as “It was cool…”, received 0.5 points when skipped.

Missing values in times did not get points. Sections where answering was stated as optional (e.g., “If you have any comments…”, “The science centre I suitable for…”), or with only one or two questions/spaces were not considered in the criteria.

Open questions with non-senses (such as an emoji or a single letter) were considered skipped. When a pilot response did not present an answer because the section was not included in the pilot, it was not considered missing.

Post-surveys were more prone to invalid data, probably because visitors come out tired or with less time available. When a response was considered invalid was decided with a cut-off value. Cut-off values were different for pre and post surveys and were decided by trial and error. Starting from a high value, the author manually checked the cases found. When the cut-off turned out to be so low that the author found it difficult to discern if some new cases found were or not invalid, the process stopped, and the immediate previous value was taken.

When a pre or post survey was deemed invalid, the full pre-post response was deleted. Two of 227 were removed from Discovery World and 22 of 1136 from Tūhura. The following table summarizes the criteria’s results.

Table
Points criteria to remove invalid responses

	Survey 1		Survey 2		Survey 3		Survey 4	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Time (2.5%)	184s		126s	57s	128s	74s	97s	51s
Time (5%)	195s		140s	78s	138s	82s	102s	57s
Time (10%)	220s		155s	100s	150s	94s	114s	67s
N skipped closed	1	1	4	3	5	4	3	3
N skipped open	0	2	0	2	0	0	0	0
N pattern	2	2	2	2	3	3	1	2
Max points	2.5	3.5	5.5	5.5	6.5	5.5	4.5	4.5
Cut-off criterion	2.5	3.5	3.5	3.5	3.0	2.5	2.0	2.0

The cut-off value is inclusive, e.g., in pre-survey 1 it would be enough to obtain 3.0 points to be a discarded response.

D.4 Holistic removal of invalid responses

There were invalid responses unable to be detected automatically by the points criteria. For example, some unreliable respondents may answer all three words, pre and post, with the same word (e.g., Fun, repeated six times). Since there is text, the points criteria Excel file cannot detect it.

They also may skip open questions or to answer them in the shortest possible way (e.g. No, idk, butterflies). Although answering the same for a whole set of a Likert-type scale can be valid, doing the same with the Questions about Knowledge (except IDK) is a lot more suspicious.

Since the unreliable behaviour tends to happen more often after the visit, the time needed to complete the post-survey also usually drops drastically in these respondents.

It is not uncommon that the unreliable respondents form part of the same group/family. In this case, it is also not uncommon that the three words are a copy-cat of each other.

Lastly, some responses coming from pilots avoided the automated Excel detector because they didn't include some items and they didn't get points from them. To determine if a suspicious pilot response was unreliable, it was necessary to go to the full pilot response (that includes other items not considered in the Excel file).

An example of a removed pilot response is a male aged 43 in Survey 4. He changed from a varied responses survey in the pre to answer uniformly all scales with the same options in the post, skipping sections and open questions and his time needed to fill out the surveys dropped from 299 to 80 seconds after the visit.

One response, out of 225, was removed holistically from Discovery World. 15 of 1114 were removed from Tūhura.

D.5 Stemming and lemmatization

Stemming was tested with an online Porter Stemmer, http://textanalysisonline.com/nltk-porter-stemmer (visited on 3 November 2018). Lemmatization was tested in R 3.4.4, library "textstem"[183]. The set of words used to compare them and the results can be seen in the following table.

[183] Packages "textstem", "koRpus", "koRpus.lang.en" and "sylly".

Table
Set of words stemmed and lemmatized

Word	Fun	Funny	Was	Education	Educational	Colourful	Interesting
Stem	Fun	Funni	Wa	Educ	Educ	Colour	Interest
Lemma	Fun	Funny	Be	Education	Educational	Colourful	Interest

Stemming struggles in some cases, like *Funny* and *Was*. It can also be confusing by reducing the words (e.g., *Educ*). Lemmatizing is clearer with word cutting (*Education* instead of *Educ*) and how it manages the root of verbs (*Be* instead of *Wa*). But it fails again in recognizing that *Funny* is related to *Fun* and *Educational* to *Education*.

Given that automated data normalization is still a long way from human capabilities and the amount of data was still manageable, the process of text normalization for the three words was as follows:

1. A set of rules was applied to the corpus.
2. Words/ideas were ordered alphabetically.
3. Lemmatization was performed manually. When in doubt, stemming was used as a guide.
4. The lemma chosen for each set of words with the same stem was the word with higher frequency[184] (e.g., if *Educational* appears 10 times, *Education* 7 times and *Educative* 8 times, all words with stem *Educ* are replaced by *Educational*).

Rules for 'the three words':

- All punctuation marks and spaces were deleted. (e.g. " Cool" → "Cool", The,Best,EVER!! → The-best-ever)
- Spelling was corrected under New Zealand spelling rules (e.g. Color → Colour, Intresting → Interesting, Neeit → Neat, Exiting → Exciting, Funnnn → Fun).
- Interjections are counted as words/ideas (e.g. Aha, Wow).
- Composed words/ideas were preferred as a single word if their form is accepted and hyphenized if it is not (e.g. The,Best,EVER!! → The-best-ever, Child Friendly → Child-friendly, Breath Taking → Breathtaking).
- Each cell was treated as a single idea/word.
- The first letter of each cell was capitalized, the rest were lowercased, unless it corresponded to a name, (e.g. Explosive! (with Amadeo) → Explosive-with-Amadeo, Ok I guess?? → Ok-I-guess).

[184] The frequency is found considering all of them, pre and post, Discovery World and Tūhura in a single list.

- Acronyms were expanded (e.g. Nz → New-Zealand).
- Composed ideas or even full sentences were treated as a single idea/word by hyphenating the components (e.g. Good For Kids → Good-for-kids, A lot of variety → A-lot-of-variety).
- Nonsenses were removed and counted as missing. The researcher double-checked that it was not a misspelling or a slang before removing them (e.g. Ga, Un).
- Composed ideas were not stemmed/lemmatized (e.g. Thought-provoking, Super-exciting).
- Contractions were expanded and hyphenated (e.g. Don't → Do-not)
- Ideas split in two or more cells were transformed into a single one (in the first place) and the rest of the cells were considered missing values (e.g. I, Don't, Know → I-do-not-know, [empty space], [empty space]).
- Words in Māori or other languages were kept in the original language (e.g. Rawe → Rawe, Scientia, Potentia, Est → Scientia-potentia-est, [empty space], [empty space]).
- Repeated words/ideas in the same cell or several cells were deleted and treated as missing data, keeping only one appearance (e.g. Very very good → Very-good, Fun, Fun, Fun → Fun, [empty space], [empty space]).
- If the three words were put in the same cell, they were split (e.g. Cool fun informative, [empty space], [empty space] → Cool, Fun, Informative).
- Emojis were removed. If the cell ended up empty, it was considered a missing value.
- No synonym replacement was done[185].
- Slang was kept in its form (e.g. Rad, Dope, Shamazing).

D.6 Nationality rules

Countries and nationalities were uniformed according to the following rules:

- USA stands for the United States (e.g. U.S.A., usa, the US)
- UK stands for the United Kingdom (including England, Northern Ireland, Scotland and Wales).
- British stands for British and English.

[185] Synonym replacement is enticing, for example Good-for-kids can be considered synonym of Child-friendly, but is it Family-oriented a synonym as well? How to draw the line where "enough" similarity ends?

- Countries and nationalities are written in the formal way (e.g., “Australian” instead of “Aussie” and “New Zealander” instead of “Kiwi”.)

D.7 Imputation

Missing values in scales were imputed under the rules already described in Chapter 2. But before, all scales were tested for consistency before imputation with Cronbach’s alpha[186]. Randomness of missing values was also calculated. Analyzing missing data is useful because it tells us if there is a pattern to study (H. Wang & Wang, 2010). Missing at random (MAR) means the missing values depend on the observed values, but not on the missing values. Missing completely at random (MCAR) is a special case where missing values do not depend on either (García et al., 2015). To find if missing data were random, Little’s Missing Completely at Random (MCAR) test was conducted. A significance below 0.05 means data were not randomly missing.

When data was not MCAR, it was holistically tested for MAR with univariates statistics (like variations in means, standard deviations and percentages of missing data), and visual assessment of missing data in search of patterns (e.g., one gender or certain age skipping, or visitors skipping an specific item).

Table

Analysis of missing data and results after imputation in Survey 3 scales

		S3-Enj	S3-FSc	S3-SCS1	S3-SCS2	S3-SF1	S3-SF2
N	Items	5	5	7	7	5	5
	Responses-T	386	386	386	386	386	386
	Item-response-T	1930	1930	2702	2702	1930	1930
	Missing-B	59	2	10	20	1	4
	IDK-B	61	0	89	42	0	0
	Missing-A	52	0	0	0	0	0
	IDK-A	15	0	41	10	0	0
Little’s test	Chi-square	47.5	1.9	157.8	127.3	4.1	8.8
	DF	44	8	112	91	4	16
	p-value	0.331	0.984	0.003	0.007	0.392	0.921
α before	Alpha	0.870	0.867	0.899	0.915	0.868	0.904
	N valid	328	384	332	342	385	382
	N excluded	58	2	54	44	1	4
α after	Alpha	0.873	0.873	0.898	0.914	0.869	0.904
	N valid	369	386	376	384	386	386
	N excluded	17	0	10	2	0	0

T stands for Total, B for before and A for after.

[186] Alpha requires not to be missing data, therefore it automatically excludes any response with even a single missing value.

Table
Analysis of missing data and results after imputation in Surveys 2 and 4 scales

		S2-SCL1	S2-SCL2	S4-MOLI1	S4-SE1	S4-SE2
N	Items	7	7	6	5	5
	Responses-T	464	464	205	249	249
	Item-response-T	3248	3248	1230	1245	1245
	Missing-B	15	20	17	16	6
	IDK-B	82	90	41	130	55
	Missing-A	7	7	13	14	1
	IDK-A	33	53	18	77	28
Little's	Chi-square	145.3	160.0	121.1	55.1	32.4
	DF	110	101	57	46	41
	p-value	0.014	0.000	0.000	0.168	0.830
α before	Alpha	0.889	0.921	0.788	0.875	0.872
	N valid	409	412	175	191	217
	N excluded	55	52	30	58	32
α after	Alpha	0.889	0.920	0.784	0.863	0.872
	N valid	455	452	198	228	242
	N excluded	9	12	7	21	7

T stands for Total, B for before and A for after.

Table
Analysis of randomness of missing data

	Univariate statistics per item							
	M_{min}	M_{max}	SD_{min}	SD_{max}	PMV_{min}	PMV_{max}	Pattern	Data
S2-SCL1	3.60	3.82	0.930	1.049	2.4	3.2	No	MAR
S2-SCL2	3.62	3.86	0.905	1.164	1.9	5.0	No*	MAR
S3-Enj	3.63	4.39	0.738	1.069	3.4	8.8	No	MCAR
S3-FSc	2.61	3.83	1.054	1.430	0.0	0.3	No	MCAR
S3-SF1	2.13	3.26	1.089	1.285	0.0	0.3	No	MCAR
S3-SF2	3.00	3.52	1.070	1.288	0.0	0.3	No	MCAR
S3-SCS1	3.30	3.58	0.964	1.183	1.6	4.7	No*	MAR
S3-SCS2	3.56	3.74	0.902	1.164	1.3	3.1	No*	MAR
S4-MOLI	3.57	4.39	0.756	1.142	2.9	8.7	No*	MAR
S4-SE1	2.14	3.06	1.279	1.504	7.2	14.1	No	MAR
S4-SE2	2.94	3.78	1.330	1.435	3.2	6.0	No	MAR

PMV means Percentage of Missing Values.
All the data marked as MCAR it is because they passed the Little's MCAR test ($p > 0.05$).
*Minimal patterns were found in these cases. For example, more post-visit visitors missing more values, probably due to tiredness. No pattern related to specific items was strong enough to consider data were MNAR.

Once it is determined that missing data are either MCAR or MAR, it is possible to proceed to imputation. Bootstraping consists in randomly choosing one element from the sample, put it back, choose another one, put it back, and so on until a new sample of the same size is formed (notice that it doesn't have the same elements, some of the original elements are repeated randomly in the resampling). This can be done a thousand times and the difference between proportions forms a normal curve. The confidence interval of this new histogram will be its mean ±1.96 SD (Efron & Tibshirani, 1994). It is considered by many as the best method as it does not depend on a formula that varies depending the situation, but applies to all equally by producing a valid normal distribution (Bland, 2015; Field, 2013;

Pezzullo, 2013). However, bootstrapping produces multiple sets and overly complicates the results.

Expectation Maximization is the method used. It produces a single set and the difference is not big with respect to bootstrapping given the small percentage of missing data.

Imputation was done only on scales of at least five items. The maximum number of missing values[187] allowed was two, regardless of if the scale had five, six or seven items. Pre and post surveys were imputed separately. When there were three or more missing values in either the pre or the post, the whole scale (pre and post) was removed and considered missing.

The item about ranking Mathematics, Explanations and Experiments is a special case. If one of the values was missing (either pre or post), it was replaced by the only possible remaining choice. For example, if Mathematics gets a two, Explanations is missing and Experiments gets a one, three is assigned to Explanations. 221 visitors were asked this question. Imputation happened 26 times in the Pre and 20 in the post.

Notice that Questions about Knowledge do not form a scale and are not imputable.

[187] Recall that IDK responses are also considered missing values.

Appendix E. Notes on Inferential Statistics

E.1 Sample size

Larger samples reduce sampling error, but do not reduce response bias (the effect of nonresponses), do not always mean a better study and the cost may surpass the benefit (Frechtling et al., 2010; Naing et al., 2006). To calculate the sample size for each survey, the following formula with finite population correction was used (Daniel & Cross, 2018):

$$n = \frac{NZ^2P(1-P)}{d^2(N-1) + Z^2P(1-P)}$$

where Z is the statistic for a level of confidence, P is the expected proportion, d is the precision and N is the population size. The Z value equivalent to the typical 95% confidence interval is 1.96. Since we cannot know P, it is usual to take the appear limit of maximum variability of the population (50%) in P=.50. The common accepted precision is d=.05 (5%).

There is no such thing as a population size of museum visitors, as the number of visitors depend on if seen per day, per month or per year. However, the formula above essentially saturates at n~384. This sample size became the ideal goal for the surveys. However, due to time constraints, only two surveys surpassed the threshold. The smallest sample is 224, which is still a good figure, as there is only a modest gain in precision when increasing the sample size above 200 (Fowler Jr, 2013).

To compare subgroups by gender and age group, a sub-sample size was calculated using the calculator provided by Soper (2019) for a two-tailed t-test under the common assumptions of d=0.5 (medium effect size), α=.05 (Type I error probability), β=.20 (Type II error probability, equivalent to .80 Statistical power) (G. M. Sullivan & Feinn, 2012). The sample size per group was 64. However, since some interesting comparisons didn't reach this target, inferential statistics were still made on sub-samples as small as 40, which would be the very minimum to throw valid statistical results[188] (Diamond et al., 2016). Any subgroup under 40 was shown descriptively, but with no inferential testing.

Acceptance/response/refusal/rejection rates are also an important consideration in calculating sample size. They were not found to be commonly reported in the literature. Among those that did, only those were surveying was clearly individual (or where the group

[188] According to (Hernández et al., 2014), the very minimum to most tests is 15, but that seems a little excessive.

was treated as a single entity) specified that they calculated it by dividing those that accepted by the total number asked. However, museum visitors usually come in group, and when surveys are to be filled out individually, that means that some in the group can accept and some can reject to participate. How to calculate an acceptance rate in such an scenario was not clearly found in the literature.

The response rate in Discovery World was determined from the number of rejections (38) and the number of pre-surveys completed (262), giving an acceptance rate of 87%. This method can be deceiving due to the unknown rate of how many individuals per group accept. To make things trickier, sometimes some of the visitors in a group want to participate, but when a dominant visitor rejects, all the group follows him/her. Also, in the case of underage visitors, in order for they to be asked for participation, parents need to agree first. It is impossible to know how many of the children of parents who rejected would have accepted.

In Tūhura a new form of calculating the response rate was used. Acceptance rate was calculated by diving the number of groups where someone accepted (456) by the number of times a group was asked (582). Tūhura's acceptance rate was 78%. Notice that this new method produces smaller rate figures than the previous one, making it more likely to include the unknown 'true' acceptance rate.

There is no agreed-on standard for a minimum acceptable response rate (Fowler Jr, 2013), but surveys that achieve a rate of 70% of higher are generally considered high quality and nonresponse is not a concern (Newcomer & Triplett, 2015). The average response rate in organizational research that collected data from individuals is 53% with a standard deviation of 20% (Baruch & Holtom, 2008).

As indicative of the effect of the response rate in uncertainty, when the response rate is 90%, then if 50% of the respondents give a particular answer, the true value is between 45 and 55%. If the response rate drops to 70%, then the uncertainty increases to 35%-65% (Fowler Jr, 2013).

E.2 Parametric testing and ordinal scales

A common debate in social sciences is if ordinal scales, such as Likert-type scales, can be analysed with parametric tests. This section discusses it.

If we want to make mathematical operations with results Likert format items, these results first need to hold the same characteristics as, at least, natural numbers.

There is an extended misunderstanding on the nature of the numbers in a Likert format. SD, D, N, A, SA usually get 1, 2, 3, 4, 5 assigned. SD, D, N, A, SA are ordered, and so are 1, 2, 3, 4, 5. But 1, 3, 4, 7, 15 are ordered as well. The big key is distance. The distance between 1 and 2 is the same as between 2 and 3 or 3 and 4. The distance is always '1', irrespective of if it is measuring centimeters, inches or seconds. To identify SD, D, N, A, SA with 1, 2, 3, 4, 5, the 'distance' between SD and D, between D and N, and so on, needs to be the same.

A number of researchers claim that the distance between each of the verbal options is the same because that's how we humans interpret it. However, there is no way to prove that all of the respondents are doing such interpretation. To explain the issue, let's consider the rulers in following images. The first image is the traditional version. The second one is a common modification where researchers remove the middle point.

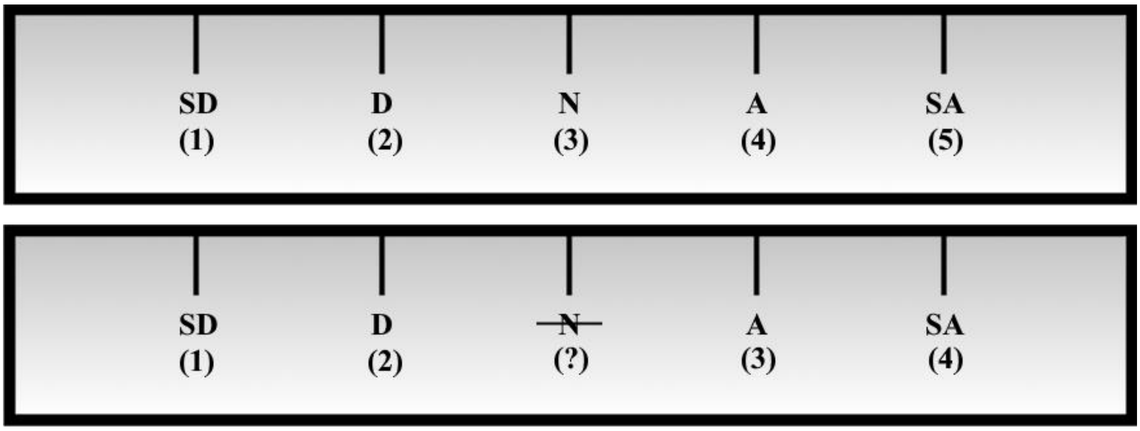


Figure. Schematic distances in Likert format with (up) and without (down) the midpoint.

What is the distance between SD and D? Both rulers seem to agree that 1. Now, what's the distance between D and A? The first ruler says 2, but the second says 1. If we want to use Likert-type scales, they need to make sense *mathematically*, but having one ruler with distance 2 and the other one with distance 1 doesn't make sense in natural numbers (even if it does ordinally).

The only article found in the literature that fully acknowledges the problem of even scales is Adelson and McCoach (2010). They compared 4-item and 5-item scales by modifying the 'distances'. For example, SD=-2 in both scales, but D=-1 for the 4-item scale and D=-0.667 for the 5-item scale. They didn't override the background problem, but they acknowledged of changing the number of options and softened the issue.

If we accept the assumption that the distance in the first scale is the same, removing one of the options completely gets rid of the validity of this assumption. We cannot simply keep modifying the rules and the rulers as we please. If we do, we automatically accept we cannot use parametric tests.

A critical assumption that needs to be fulfilled for parametric testing is normality, but normality is a characteristic of interval scales, not ordinal scales[189]. Not understanding these statistics leads to many researchers choosing tests arbitrarily with no justification. For an excellent discussion on the issue, see Clason and Dormody (1994).

Even if we decide to keep only the 5-point version. What is the average of Disagree and Agree? Is it Neither? Moreover, what is the sum?

When we assign numbers to a scale, *they are not figures, they are symbols*. They just help us to quickly visually the results. While we can do mathematical operations on numbers, we cannot do the same on symbols. It would be like averaging X and Z, is it Y?

Does it mean that parametric analysis should be banned from Likert-type scales? Not quite so. We can still justify treat Likert-type scales as interval under certain circumstances. Before anything, it should be understood that it is an approximation and that it is based on assumptions.

First, let's not remove the middle point. Second, a critical assumption is that most people unconsciously give the same distance to points in a Likert-type scale. Third, if there is variation in the second assumption, they don´t have preferred direction of variation. Therefore, if we have 'enough' data, these variations would tend to cancel out. There is no mathematical way to know how much data is enough, but we do know that the more items and the larger the sample, the more probable it is that these variations cancel out.

If we accept these assumptions it is plausible to proceed to consider these data as interval. It is not still enough to use parametric tests directly, but it allows to test for parametric requirements, such as normality.

E.3 The meaning of *p*

p is one of those statistical things that tend to be misinterpreted quite frequently. A p-value is *not* the probability of the null hypothesis being true ($p(H_0|D)$). It is, given that the null hypothesis is true, the probability of getting the observed data ($p(D|H_0)$) (Gigerenzer, 2004; Haller & Krauss, 2002). In other words, p is the probability of getting the observed result, but given as condition that the hypothesis is true. It does not apply if it is not true. *The*

[189] It doesn't matter if a bunch of number-symbols can be plotted in a nice Bell curve, symbols cannot be normal, it is an ontological problem beyond the shape.

hypothesis being true is a pre-condition. Thinking that the opposite holds, that p gives the probability of the null hypothesis being true is known as the 'inverse probability fallacy' (Cumming, 2013).

Sometimes the p-value is compared to the α value (probability of making a Type I error). But, actually, α is set beforehand, usually at .05 (or .01). If p ≤ α, then the null hypothesis is rejected.

E.4 Effect size and confidence interval

Effect size (ES) and confidence interval (CI) were already thoroughly explained in Chapter 2. This section drills down a bit more in what they mean and how they are calculated.

CI has two main assumptions: normality and random sampling. If one of them is not fulfilled, a different approach might be better (Cumming, 2013). However, for large samples, normality can be overridden by the central limit theorem (Cumming, 2013).

The standard error (SE) is the standard deviation (SD) of the sampling distribution of the sample mean (Cumming, 2013). In other words, if we have a population, the sample's mean will not necessarily coincide with the population's mean. Sample's SD would describe how spread the sample's data are, but nothing about the population. If we take lots of samples, their means form another distribution (normal, according to the central limit theorem). The SD of this new distribution is the SE, and SE, indeed, is inferential information that tells us, based on the sample data, how confident can we be that the sample's mean coincides with the population's mean.

If we want to be 95% certain that we got the true mean of the population, we need to consider a range around the sample's mean in what is called Confidence Interval (CI). Numerically, $95\%\ CI = [M - 1.96 \times SE, M + 1.96 \times SE]$, where $SE = {}^{\sigma}/_{\sqrt{N}}$ and $\sigma \equiv$ [190] $SD_{sample} = \sqrt{(\sum X_i - M)^2 / _{N-1}}$ (Cumming, 2013).

Another tool to know about the results is the effect size (ES), that adds the 'how much' (size) to the effect we are interesting in (Cumming, 2013). It is not something unique, there are many ways to represent the size of an effect. For example, a mean, a median, the difference between two means, the correlation coefficient, etc. (Cumming, 2013).

[190] Sometimes this is called "sample SD", and replacing N-1 by N is the "population SD". This is not exactly so, SD with N-1 is the inferential SD, it infers features of the population based on the sample. Replacing N-1 by N would give us a descriptive SD, i.e., it describes the data alone, and that may be adequate for some samples if the goal is to describe the sample, not the population it was taking from.

The fourfold point correlation coefficient (ϕ) is frequently reported in 2x2 contingency table in behavioural science, but actually is identical, i.e., $w = \phi$ (Cohen, 1988).

When the contingency table is larger than 2x2, the effect size becomes Cramer's V (a.k.a. Cramér's V or Cramér's ϕ'). In this case, how to measure if it is small, medium of large, depends on the degrees of freedom, but the transformation is quite simple

$$w = \phi' \sqrt{r-1}$$

where r is the number of rows in the r x k contingency table[191]. Note than when the table is 2x2, w becomes ϕ. Since 3x3 is a common table used in this thesis, the case of ϕ' for r=3 (Cramer's V for 3x3).

Eta squared (η^2) is equivalent to r^2 (Tomczak & Tomaczak, 2014) and it can be transformed into the effect size index f (Cohen, 1988):

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}}$$

Several effect sizes can be transformed among them, as between d and r is (Ferguson, 2009):

$$d = \frac{2r}{\sqrt{1-r^2}}$$

E.5 Small, medium and large effect size: beware of wrong values

To decide if an ES is small, medium or large, the most common reference is J. Cohen (1988). But there is a lot of misinformation around citing Cohen incorrectly. Examples of misuse are Lenhard and Lenhard (2016), G. M. Sullivan and Feinn (2012) and Ferguson (2009).

In the first example, Lenhard and Lenhard (2016) cite Cohen as reference to say that [0.0,0.2) is "No effect", [0.2,0.5) is "Small effect", [0.5,0.8) is "medium effect" and >=0.8 is "large effect". But such intervals are never mentioned by Cohen. In fact, Cohen, in 'Illustrative Example 2.6' gets d=0.4 and he himself classifies it as "small to medium value" (Cohen, 1988, p. 50).

[191] Do not confuse this *r* with the correlation coefficient.

Cohen’s values for small, medium and large are not lower values, they are the values ‘around’ which small, medium and large effects can be considered (Cohen, 1988, 1992; Durlak, 2009; Field, 2013).

The erroneous interpretation is quite spread. G. M. Sullivan and Feinn (2012) has been dangerously widely cited. In their article, they mention that the common values are d=0.2, 0.5 and 0.8 and 1.3 for small, medium, large and very large respectively. And then Pearson’s r as 0.2, 0.5 and 0.8 for small, medium and large and mention that the table is adapted from Ferguson (2009). First of all, Ferguson does not mention 1.3 as very large anywhere. Second, Ferguson mentions that Cohen suggests r=0.3 and d=0.5 as cut-off for moderate effects, this time citing J. Cohen (1992). So far, we have two misreports by Sullivan and Feinn. But if we go to J. Cohen (1992), it turns out that the article simply ratifies that small, medium and large are for d=0.2, d=0.5, d=0.8, as in the original book of 1988. It doesn’t mention they anything like a lower cut-off either. Also, the equivalent values for small, medium and large are r=0.1, r=0.3, r=0.5 (Cohen, 1988, 1992).

Ferguson (2009) also recommends a minimum effect size (RMPE) of 2.0 for Odds Ratio; 3.0 would be moderate effect and 4.0 strong effect. However, finding the equivalent RMPE for small effect (from values of d), it would be 0.41, moderate effect would be 1.15 and strong effect would be 2.70. Not coinciding with his own recommendations.

Both references, G. M. Sullivan and Feinn (2012) and Ferguson (2009) not only mis-cite Cohen, but add new values out of the blue, referencing other authors that never mentioned them.

E.6 Confidence intervals of effect sizes

In Chapter 3 it was explained how the confidence interval for Cohen’s d was calculated. Although Cohen’s d was the only effect size whose confidence interval was calculated, some common effect sizes’ confidence interval can be found as follows.

r^2 (Cohen, Cohen, West, & Aiken, 2014)[192]:

$$95\%\ CI\Delta = R^2 \pm 1.96\sqrt{\frac{4R^2(1-R^2)^2(n-k-1)^2}{(n^2-1)(n+3)}}$$

[192] r is usually written in lowercase in bivariate cases, while R is used in the context of multiple correlations (i.e., more than one predictor). However, when r is squared, the common practice is to represent it as R2 as well (Field, 2013).

where n is the sample size (larger than 60) and k is the number of independent regressors (in this thesis all cases are k=1).

r (Field, 2013). In this case the process is not as direct as before. First, it is necessary to transform r into z-score (to make the distribution normal):

$$z_r = \frac{1}{2} log_e \left(\frac{1+r}{1-r} \right)$$

The resulting value has a standard error of:

$$SE_{z_r} = \frac{1}{\sqrt{N-3}}$$

The CI boundaries in the z space are therefore:

$$CI = \left(z_r - 1.96 \times SE_{Z_r}, z_r + 1.96 \times SE_{Z_r} \right)$$

And these values can be transformed back from the z space as:

$$r = \frac{e^{2z_r} - 1}{e^{2z_r} + 1}$$

Due to the transformations between spaces, these values are not symmetrical. Notice that this formula only works for tests with one group. For example, it cannot be applied to Mann-Whitney tests.

Technically, it is possible to transform an effect size from one format to Cohen’s d, making interpretation and calculation of effect size’s confidence intervals homogeneous. For example, the transformation from R^2 to r is straightforward (root squared). From r to d is as follows (Borenstein et al., 2009; Cohen, 1988):

$$d = \frac{2r}{\sqrt{1-r^2}}$$

The Odds Ratio can also be transformed into d (Borenstein et al., 2009):

$$d = LogOddsRatio \frac{\sqrt{3}}{\pi}$$

However, there are restrictions for which it was decided not to do these transformations. From OR to d “It assumes that an underlying continuous trait exists and has a logistic distribution (which is similar to a normal distribution) in each group. In practice, it will be difficult to test this assumption” (Borenstein et al., 2009, p. 47).

From r to d it is a lot more complicated in terms of interpretation. First, it is only valid when both groups are equally numerous. Second, the interpretation of r is .1 small, .3 medium and .5 large. But taking r=.5 (large) and transforming into d, gives d=1.155, which is an extremely large value, way more than the expected 0.8. J. Cohen (1988) gives a comprehensive and mathematical explanation of why this happens and how to interpret it. The explanation goes beyond the scope of the appendix.

Appendix F. Coding

F.1 Coding manuals for surveys' open questions

Appearances of categories were counted dichotomously. In other words, it is not about how many times the category was mentioned, only whether it was mentioned or not. A comment can include more than one category unless the coding manual specifies otherwise.

Coding manual A1: "It was cool learning about…" [by exhibit - Discovery World]

This code was used in "It was cool learning about…" (Survey 1). More than a code, this manual was a guide for counting the number of times each exhibit was mentioned. Categories OELE and OE were mutually exclusive.

Table
Coding manual A1

Category	Definition	Inclusions	Exclusions
Tropical Forest (TF)	Mentions related to the Tropical Forest.	"the butteflies are beautiful and the chickens are cute." "The rainfall levels". "Turtles".	
Planetarium (Pl)	Mentions to the shows and stars.	"The stars and other things".	
The Science Show (SS)	Comments that mention the science show either explicitly or indirectly through the topics it was about those weeks (fire, frozen marshmallows, sound, explosions, nitrogen, dry ice).	"Explosive bubbles". "Frozen marshmallows". "Sound waves".	"Electromagnetism.... okay. It was cool to learn that i can sort of called a banana a gamma ray gun...." (sounds like something Amadeo would say, but it's on a panel in the Light Zone that is not associated to any exhibit, it goes to NEM).
Mind Ball (MB)	Mentions of the mind game and its characteristics, like theta waves and relaxation.	"Brain waves". "Thetherma waves was pretty cool" (it's likely they tried to spell "The theta waves…", the other option is The Theremin, but that exhibit does not mention waves).	
Plasma Room (PR)	Mentions of the plasma room or these exhibits: 1) Jacobs ladder, 2) Neon lightning, 3) Plasma plates, 4) Plasma tube.	"Plasma". "Electric lightning stuff" (Neon lightning).	
Dung Beetle (DB)	Mentions of the Dung Beetle (Push Poo is the official name) exhibit.	"That a beetle pushes 1000 times its weight".	

Earthquakes (EQ)	Mentions of the Earthquake exhibit.	“Different buildings vibrating in different strengths of earthquakes.” “Building resonance”.	
Other Exhibits related to Light and Electromagnetism (OELE)	Mentions of Light Zone exhibits, except those in the Plasma Room (only include optical illusions if the explanation is physical, not mind related): 1) Coloured shadows, 2) Frozen shadows, 3) Funhouse mirrors, 4) Hologram, 5) Paint with light, 6) Ray table, 7) The Theremin, 8) Hand shake, 9) Electromagnetic crane, 10) Infinite pipe (official name: How long is this pipe), 11) Kaleidoscope. Also include those where it is clear they come from the Light Zone, even if it is not clear which one (like “Light”)	“Phosphorescent materials and photos” (Frozen shadows panel talks about phosphorus and photography, this is only one exhibit mentioned). “Light refraction!” (Ray table). “Lazers” (Ray table is the only one with lasers).	“Lights, and the way you can use for example electricity and I also thought it was cool seeing all the different things you can do” (it is clear it’s related either to OELE or MG, but it is not possible to associate to a specific exhibit, it goes to NEM). “The optical illusions and how they trick your eyes” (this is mentioned in the Thatcher illusion, which has not a physics explanation, it goes to OE).
Other Exhibits (OE)	Mentions of exhibits that do not belong to any of the previous ones: 1) Air cannon, 2) Cloud machine, 3) Concrete plank (official name: Walk the plank), 4) Hover ball, 5) Icy bodies, 6) Kinetic sculpture (official name: Start the ball rolling), 7) Slide, 8) Weather station, 9) Air hockey, 10) Colour me, 11) Foosball table (official name: Table soccer), 12) Gear-shaped table, 13) Impress me, 14) Piano (official name: Foot stomping), 15) Reaction timer, 16) Smell this, 17) Squared-shaped table, 18) Torsion wave, 19) Thatcher Illusion, 20) Stroop effect, 21) How many nails?, 22) Odd ball*. Also includes those where it seems there is an exhibit involved and it’s not clear which one, but definitely cannot fit in any of the other categories.	“Planets. Butterflys air cannon nails” (“air cannon” is an exhibit, so “How many nails?” is, therefore, the count is 2 OE, note that it also ticks off Planetarium and Tropical Forest).	“How sound waves travel” (sound is one of the Science Show topics). “Forces” (there are several exhibits related to forces, from the dung beetle to the electromagnetic crank, it goes to NEM).

No Exhibits Mentioned (NEM)	Tick this off if none exhibit was mentioned in the comment, or when it seems there is an exhibit involved, but it could fit in more than one of the categories above.	"Everything!" (can't relate with a specific exhibit). "Sinis" (probably tried to spell "science"). "how kids can enjoy learning".	

*Odd ball is an explanation of why the Hover Ball is suspended in the air thanks to a difference in pressures in the air. They both are the same physical exhibit. Only count them twice if they specifically mention both, otherwise, it is only one. According to me, they were never explicitly mentioned twice.
Tricky example: "Butterfly lifecycle brainwaves and relaxation sound and force". It is not "brainwaves" and "relaxation sound", it is "brainwaves relaxation" and "sound" (however, both would correspond to MG and SS).

Coding manual A2: "It was cool learning about…" [by exhibit - Tūhura]

This code was used in "It was cool learning about…" (Survey 2), "Can you give an example of what you learnt?" (Survey 3) and "Have you say! Let us know what you liked of didn't like of Tūhura's exhibits and how we can improve them" (Survey 4). Similarly to the previous one, this manual was aimed at guiding how to count the number of times each Tūhura exhibit was mentioned. Categories with an * means that more than one exhibit is included in the same category, but it was usually too difficult to know which one they were talking about. If a visitor explicitly mentions both exhibits, it was counted twice. E.g. "How dams work and looking at the topography exhibit" (code 'Topography' was counted twice, as the comment refers to two exhibits).

Table
Coding manual A2

Category	Definition	Inclusions	Exclusions
Magnetic Sand+ (MS)	This exhibit is a bowl with black magnetic sand and two powerful magnets. The official name is Magnetic Sculptures.	"I liked the learning about the black sand."	"Sand displays" (Magnetic sand always involved the magnets or iron sand or black sand; sand alone is Topography).
Floating in Copper (FC)	Mentions of magnetism or magnets need to include some reference to copper, or floating, metal ball, tube, to be considered related to Floating in Copper.	"Magnetic forces like the black sand and the floating magnetic ball".	
Magnetism+ (M)	MS and FC are both related to magnets. It is expected that almost all comments related to magnetism talk of Magnetic Sand, which is extremely popular, while, on the other side	"Magnets".	

	of the spectrum, Floating in Copper (FC) is not popular. However, if it is not clear which one the comment is referring to, keep it in this special code.		
The Void (TV)	The Void is an exhibit that talks about the big bang from a Maori perspective and includes “infinite” lights in a dark room with water sounds in the background.	“Don't know there were lots of activities but mine would be 'the void'”. “Big bang representation”.	“Dark matter” (This was a show at the planetarium, it goes to NEM).
Topography* (To)	Topography and Land and Water are two exhibits about topography and landscape shaping. It would be sometimes too difficult to know which one visitors are referring to, since both use some sort of sand. So, both are coded together. Topography uses sand and a projector to virtually project a landscape and “make rain”. Land and Water is a big table where water is real, flowing like in a river and visitors build dams.	“The water”. “Erosion”.	“The black sand” (the Magnetic Sand is black).
Sound Bite (SB)	In Sound Bite you bite a straw and hear music trough the bones of your head.	“Sound travelling through solids.” “You can hear stuff while eating plastic straws”.	
The Buttons (TB)	In the Red/Blue buttons you just press them as fast as possible.	“The void and the butterflies But... the buttons were the b b b b b b b b b b BEST”.	
Mood Ball (MB)	The Mood Ball is a ball that goes over your head to block the exterior. The only you see is colours changing, which changes your mood.	“The colour stuff and how it efects your moods”.	
Rocks Table (RT)	A table with rocks.	“About the different rocks”.	
Monochromatic Room (MR)	Monochromatic Room has a sodium lamp that emits only one wavelength and everything looks yellow/brown-ish. Using a torch, it is possible to discover all the colours actually present in the room.	“The yellow room” (the room looks yellow-ish). “The colour spectrum and how it changed and also about torque force and how they get the butterfly's in” (in the Monochromatic Room, colours change if we add the full spectrum of light with the torch;	

		the other category is the Torque Table). "The electricity stuff and when you put your hand to the circle it will produce electricity through you. The street light one was interesting" (the first one is Plasma, the second part is MR). "How colorblind see" (it is not actually how a colour blinded person would see, but that's how they interpreted the exhibit).	
Animation Station (AS)	You can make a mini film frame by frame at the animation station.	"The void The flight zone Butterflies Souund waves Animation" (This one has TV, FZ, SB, AS).	
Plasma* (P)	There are two exhibits that talk about plasma, States of Matter and Plasma: the fourth state. In Discovery World they were part of the Plasma Room (Plasma Plates and Plasma Tube). Both are coded together.	"Getting shocked" (Plasma Tube produces little electrical shocks). "States of matter". "Light and electricity" (electricity refers to the electrical shocks of the plasma exhibit).	
Skeleton Bike + Human Body (S+H)*	Skeleton Bike is a skeleton riding a bike, Human Body is a 3D model of the body, but sometimes it is difficult to know which one they are talking about. Bike, Skeleton and Human body are direct reference to this one. In the Biozone there are bones, but they are skulls.	"Butterflies, skeletons". "How your body works".	
Newton's Prism (NP)	Some terms related to this exhibit are Newton, prism, refraction, dispersion, separation into colours. Wavelength as well, but it can refer to other exhibits.	"Light refraction".	"Color mixes" (Newton's Prism is about splitting light into colours, not the opposite, this is CS).
Earthquakes (EQ)	Earthquakes	"Earthquakes".	
Gravity Wall + Vacuum Drop (G+V)*	Gravity wall is an exhibit where a ball rolls down. Vacuum Drop removes the air from inside a tube to see how a feather and a ball fall at the same speed. Unrelated, but it is hard to say if "gravity" refers to one or the other, so they are kept together.	"Gravity". "vacuums".	

Dancing with Light (DL)	At Dancing with Light you move and a Kinect scans you and projects your image on a screen, but distorted, delayed, in colourful layers.	"The world of light, the dance experience, and sensory experiences that are appealing and enlightening to adults as well as children." (the only valid category here is DWL).	
Chicken Embryos (CE)	Chickens, Embryos and Eggs are key words to identify this one.	"The fertilisation of the egg" "Chicken fertilised egg, great white" (the first part is CE, but "great white" is not the white of the egg, it is the Great White Shark, from the planetarium; since it does not add anything, CE is the only category to tick off).	
Chicken Wire (CW)	This one is about sensory. You rub your hands on chicken wire and it feels like velvet. Do not confuse with CE.	"The way things respond to human touch and changing common perception."	
Infrared Camera (IR)	Infrared camera to detect the temperature of the bodies.	"Infrared light".	"Light" (it doesn't mention that it is infrared, goes to NEM).
Tornado (T)	The tornado is created with water and it is possible to send a 2D laser to see the eye of the tornado.	"Light,volcanoes,tornados".	
Flight Zone (FZ)	At the Flight Zone you create helicopters with paper and paper cups and they fly.	"Flight zone". "Helicopters, maps,stop motion, ultraviolet light,a d most of theother exhibits" (the helicopters plus several others).	
Torque Table (TT)	The Torque Table is a turning disco where you can roll objects over and see what happens.	"Torque! And butterflies." "Light and Momentum" (momentum is a term for objects moving, in this case, circular momentum).	
Coloured Shadows (CS)	This one shows that mixing Red, Blue and Green lights you can make white light.	"Color mixes".	"How the mirrors and coloured lights made you feel in the void" (goes to The Void).
Volcanoes (V)	Volcanoes.	"Volcanoes and waterways" (this one includes Volcanoes and Topography).	
Ultraviolet Camera (UV)	An ultraviolet camera that shows yourself on a screen.	"Helicopters, maps,stop motion, ultraviolet light,a d most of theother exhibits".	
Fun House Mirrors (FHM)	Funhouse mirrors from Discovery World that now are before the Tropical Forest.	"mirrors".	
The Slide (TS)	The DNA slide.	"You go faster without having shoes onin the slide".	
Seismometer (S)	An actual seismometer.	"Watching the live seismometer was amazing. Fascinating to see how the earth is active."	
Mimosa Plant (MP)	This plant closes its leaves if touched.	"sensitive plants".	

Microscopes (Mic)	They are microscopes.	"Te puaka matariki and how gross skin looks under microscope" (Te puaka matariki is the Maori new year, a show in the planetary, not counted here, only Microscopes).	
Other Exhibits (OE)	I expect to have covered all the exhibits mentioned in the comments. If you considered there is one that does not fit in any of the above, tick it off and record the ID number.		"Humans are more dangerous to sharks than to people." (Sharks was a show at the planetarium those weeks, it goes to NEM). "Animals" (it' related to the Biozone, but it doesn't specify the exhibit, goes to NEM).
No Exhibits Mentioned (NEM)	Tick this off if none exhibit was mentioned in the comment.	"Light and touching things" (no specific exhibit). "Games". "Lights". "Mind tricking". "Reflective lighting" (related to the light zone, but who knows what exhibit).	

Coding manual B1: "It was cool learning about…" [by topic - Discovery World]

This code was used in "It was cool learning about…" (Survey 1). The difference with A1 is that this time it was a formal code to detect the topic mentioned.

Table
Coding manual B1

Category	Definition	Inclusions	Exclusions
Tropical Forest (TF)	Mentions related to the Tropical Forest.	-"the butteflies are beautiful and the chickens are cute." -"How much silk it takes to make a tie" -"The rainfall levels -"Turtles"	
Light and Electromagnetism (LE)	Anything related to light, vision, electricity, magnetism, electromagnetism, lightning, plasma room (only include optical illusions if the explanation is physical, not mind related). Use PR and OELE from Code A as reference.	-"Phosphorescent materials and photos" -"light and prisms" (note that they are counted only once) -"The magnetic waves"	-"Southern lights" (this comes from the planetarium, Pl)
The Science Show (SS)	Comments that mention the science show either explicitly or indirectly through the topics it was about those weeks (fire,	-"Explosive bubbles" -"Frozen marshmallows" -"Sound waves"	-"It was cool to learn that i can sort of called a banana a gamma ray gun...." (this is mentioned on

	frozen marshmallows, sound, explosions, nitrogen, dry ice).		a panel in the plasma room, it goes to LE)
Non-Light and Electromagnetism (NLE)	Mentions of topics not related to LE, but that come from the exhibits zone, it does not include SS, TF and Pl. Use MG, EQ, DB, OE from Code A as a reference.	-"Brain waves" -"The clouds, bendable concrete" (note that they are two topics, but counted once). -"Thetherma waves was pretty cool" (it is likely they tried to spell "The theta waves…")	
Science in General (SG)	Nothing specific, when they liked everything or just science in general.	-"Everything!" -"All"	-"how your imagination works" (it is specific, but doesn't fit in any category, it goes to OT) -"other things" (it is not specific, and it doesn't talk about the science in general, goes to Other)
Planetarium (Pl)	Mentions of the planetarium and the space.	-"The stars"	
Other (Ot)	Other comments that do not fit in any of the above.	-"How kids can enjoy learning"	-"The sinince" (it was a young respondent, she was likely trying to spell out "The science", goes to SG) -"How sound waves travel" (this is a topic of the SS)

Tricky example: "Butterfly lifecycle brainwaves and relaxation sound and force". It is not "brainwaves" and "relaxation sound" (which could be interpreted as MG and Other), it is "brainwaves relaxation" and "sound" (which corresponds to MG and SS).

Coding manual B2: "It was cool learning about…" [by topic - Tūhura]

This code was used in "It was cool learning about…" (Survey 2), "Can you give an example of what you learnt?" (Survey 3) and "Have you say! Let us know what you liked of didn't like of Tūhura's exhibits and how we can improve them" (Survey 4). The difference with A2 is that this time it is a proper code that detects the topic mentioned.

Notice the coding topic does not necessarily coincide with official Tūhura zones. For example, the code's category of *Torque Table* is Mechanics, although it officially belongs to the Unseen Forces Zone. General Light (GL) and Light and Electromagnetism in physics (LE) are mutually exclusive. Only select GL if LE is not selected.

Dancing with Lights is technically an exhibit about Light, but visitors mainly consider it in terms of an artistic experience, and therefore it was placed in Psychology, Art and Culture.

The *DNA Slide* is technically part of the Biology section, but visitors take it as a regular slide and it was deemed more appropriate to Mechanics.

Table
Coding manual B2

Category	Definition	Inclusions	Exclusions
Tropical Forest (TF)	Mentions related to the Tropical Forest. Note that it does not includes "insects" and arthropods.	"Butterflies. A real treat to see them". "Tropical rain forest".	"It was quite interesting learning about new speices concedring i didnt know a lot about them" (there are different species in the Bio Zone, not clear if it is TF or Bio, then it goes to Ot). "Insects" (this one goes to Bio).
General Light	All comments that simply mention "light" and it is not possible to distinguish if they are talking about lights in The Void or Dancing with Light (PAC) or from a physics perspective (LE).	"Light". "The world of light". "Light and momentum" (GE and M).	"Light the monochrome room was very cool." (this is part of LE). "Light and shadows" (this is the "Coloured shadows" exhibit, it is LE). " Light and electricity" (not clear what exhibit, but electricity is LE and that's it, GL cannot be selected).
Light and Electromagnetism in physics (LE)	Anything related to light, vision, electricity, magnetism, electromagnetism, lightning. All from physics perspective. The following exhibits belong to this category: 1) Magnetic Sand, 2) Floating in Copper, 3) Monochromatic Room, 4) Plasma, 5) Newton's Prism, 6) Infrared Camera, 7) Coloured Shadows, 8) Ultraviolet Camera, 9) Funhouse Mirrors.	"Lights and electromagnetic waves" (counted only once, it is from the physics perspective). "Black sand" (it is Magnetic Sand).	"Big bang representation" (it is The Void, goes to PAC). "How the mirrors and coloured lights made you feel in the void" (it specifies it is The Void). "Sand displays" (plural, no mention to black or magnetic, it is the topography exhibits, ES). "Physics" (there is no category for physics exclusively, it goes to Ot).
Biology (Bio)	Anything related to biology, including the BioZone or the human body, insects and artropods (except butterflies and tarantulas and caterpillars). The following exhibits are related to this: 1) Skeleton	"The diffeent thing about the human boys and bugs bodys also the diffrent kinds of butterflies and spiders,caccons and turttles." (Bio and TF).	"Sharks, how fast things move" (it is the planetarium show).

	Bike, 2) Human Body, 3) Chicken Embryo, 4) Mimosa Plant, 5) Microscopes, 6) Soundbite, 7) Collection of animal and human skulls and bones.	"The night sky and experiencing the tropical rain forest, the marine room was also interesting and people through the ages" (Pl, TF, Bio). "Senses".	
Psychology, Art and Culture (PAC)	Topics where psychology, art, culture, history, are the thread or where perception of the world is involved: 1) The Red/Blue Buttons, 2) Mood Ball, 3) Chicken Wire, 4) The Void, 5) Animation Station, 6) Dancing with Light.	"Different things. Working together to do pac mac. Topography" (PAC and ES). "Science how things work learning more about are cultural and fun for kids" (PAC and SG). "The velvet hands activity i didnt know what it would feel like". "Science and history" (SG and PAC).	"How scared people are of large spiders. How most of my group didn't remember much about light." (the fear to spiders come from Tropical Forest, the other one is LE).
Earth Sciences (ES)	Topics related to the Earth, like the following exhibits: 1) Topography, 2) Land and Water, 3) Rocks Table, 4) Earthquakes, 5) Tornado, 6) Volcanoes, 7) Seismometer.	"Butterflies Earthquake structure Erosion" (TF and ES, counted once). "About the different rocks". "Tornados".	"Seasons in NZ" (it is mentioned in the Tropical Forest).
Mechanics (M)	Topics where dynamics and cinematic are involved, like in: 1) Gravity Wall, 2) Vacuum Drop, 3) Flight Zone, 4) Torque Table, 5) The DNA Slide.	"Gravity". "Torque! And butterflies." (M and TF).	"Dark matter" (it is a show at the planetarium).
Science in General (SG)	When they expressed they liked learning something, but not expressed what specifically.	"Test physics concepts. Butterflies watch fly." (SG and TF). "How things work". "Sciencia". "All". "Science and the human body" (SG and Bio). "Having hands on learning opportunities" (note that "learning" makes it fit here).	"All the different ways our bodies interact in the world around us" (Close to SG, but bodies interacting sounds more like Biology). "touching things" (The presence of Science or learning is not clear, it goes to Ot).
Planetarium (Pl)	Mentions of the planetarium and its shows.	"Constellations and galaxies". "Humans are more dangerous to sharks than to people."	
Other (Ot)	Other comments that do not fit in any of the above.	"Games". "Stuff".	"Butterflies and exbets" (Butterflies is TF, and "exbets" probably is a misspelling of "exhibits", that goes to SG).

Coding manual C1: “It would be cool if Discovery World had an exhibit about…”

This code was used in “It would be cool if Discovery World had an exhibit about…” (Survey 1).

Table
Coding manual C1

Category	Definition	Inclusions	Exclusions
Animals (A)	All types of animals, as long as they currently exist.	“More cool fish”. “Snakes”. “Tigers”.	“Animal evolution” (it fits better under Biological Systems).
Extinct Animals (EA)	Extint animals.	“Dinasours”.	
Biological Systems (BS)	From genetics to ecosystems, all biological systems that refer to its scientific study, not to animals or tropical forest	“Native ecosystems”. “Plants”. “Anatomy”. “The human body, digestion, breathing, brain function” (note they are several things, but counted only once because they all belong to the same category).	
Light and Electromagnetism (LE)	Anything related to light, vision, electricity, magnetism and electromagnetism.	“Magnetism”. “More illusion” (this might be physical or physcological, but will be kept in LE, as it is related to vision).	
Earth Sciences (ES)	Things related to this planet and its study, except auroras.	“Earthquakes”. “How different minerals were made, coal, diamonds, iron etc”. “Climate change”.	“gravity” (There is gravity on Earth, but not only here, it fits better in Forces and Mechanics, as it is a force).
Universe (U)	Mentions to the outer space. Including auroras (they are a terrestrial phenomenon, but with space origin that people identify more with night sky/space	“More space things possibly”. “Auroras” (they technically happen in the Earth, but with extraterrestrial origin and more identified with the space).	
Sound (S)	Mentions to sound.	“Echoing sounds”.	
Forces and Mechanics (FM)	Mentions to forces and mechanical systems or mechanics are a branch of physics.	“Mechanics”.	
Technology (T)	Comments related to technology (applications).	“Clocks/gears or electronic circuits”.	“Nuclear power” (the technology to harvest it is not mentioned, it goes to Other). “Video games” (it is not about the technology behind them, it goes to Other).

Science in General (SG)	Nothing specific, when they want anything, or just science in general.	"More nz science things" (It is specifically NZ science, but still very general).	
Good as is (GAI)	If they consider all they want is covered, it goes here.	"Hard to fault what u have already Y" "Everythings good".	
Avoiding Comment (AC)	Technically they are comments, but they do not say anything, it was just to fill something in or they don't know what to say.	"Unsure".	"Anything" (it is not really avoiding to comment, it is more or the side of add something, anything! It goes in Science in General).
Other (Ot)	Any comment that does not fit in any of the above categories.	"Gas:fart!" "Fireworks".	"Something interactive that followed up from the planetarium shows." (they have shows and many things, but mainly related to Universe).

Tricky example: "Animal phsycology" does not go to AD (Animal), but to Other (because psychology is more specific than animal).

Coding manual C2: "It would be cool if Tūhura had an exhibit about…"

This code was used in "It would be cool if Tūhura had an exhibit about…" (Survey 2).

Table
Coding manual C2

Category	Definition	Inclusions	Exclusions
Animals (A)	Mentions to animals that currently exist on Earth.	"Carnivores plants". "Cats. How there purr can affect things." "Sea creatures coral etc". "Tuatara".	"The ocean" (they are not mentioning sea animals, it goes to ES). "Horses evolving" (evolution is a heavier concept and it fits better in BS).
Extinct Animals (EA)	Mentions to animals that already went extinct.	"Dinosaurs". "Moa".	
Biological Systems (BS)	From genetics to ecosystems, all biological systems that refer to its scientific study, not to animals or tropical forest. It includes the human body and brain.	"That old jedi mind ball kit that would require you to move the ball to the oppositions side via brain activity" (it is related to brain activity). "Facial movements".	"Biomechanics = forces and human motion. More of these concepts could be explained with the skeleton on the bike." (Biomechanics is closer to Technology).
Physics and Chemistry (PC)	These two fields, including light and electromagnetism, forces, mechanics and sound.	"Heaps of black sand" (they are talking of the magnetic sand). "Friction is so good". "Heat".	"Zero gravity" (it fits better in Universe). "Laser tag" (it is the game, not about laser, it goes to Other). "Aerodynamics" (it is physics, but they are

			likely referring it from a technological point of view, goes to TT)
Earth Sciences (ES)	Things related to this planet and its study.	"Earth quakes". "I also enjoyed the upstairs geology landscape section; could do more to link them". "Water".	"Earth spinning" (it has a direct effect on Earth, but seeing it spinning can only be seen from the outer space, fits better in U).
Universe (U)	Mentions to the outer space.	"Cosmo". "Planets". "Gravity".	"Rocket one" (Rockets fit better in TT).
Transport and Technology (TT)	Comments related to means of transport and technological applications.	"Computing turing machine. Model chemistry reactions model cell operation. Model control systems" (T and PC). "Something about planes or streamlining cars / vehicles". "Artificial hearts".	
Science in General (SG)	Nothing specific, when they want anything, or just science in general.	"Anything".	
Good As Is (GAI)	If they consider all they want is covered, it goes here.	"It was enough".	
Avoiding Comment (AC)	Technically they are comments, but they do not say anything, it was just to fill something in or they don't know what to say.	"Not sure". "Dont know".	"Unsure ‚Just about covering it all" (fits better in GAI).
Other (Ot)	Any comment that does not fit in any of the above categories.	-"The tube was fantastic!" -"Cricket and soccer" -"More things like the void" -"School groups would enjoy. More physical activites." -"Metal"	-"Aliens" (it goes to Universe)

Coding manual D1: Comments - Discovery World

Comments come from Survey 1. In this case, the Comments section appeared before and after the visit. Pre and post comments are merged into a single comment to avoid duplications of category mentions.

Table

Coding manual D1

Category	Definition	Inclusions	Exclusions
Enjoyment / Appreciation (EA)	Comments where the visitor explicitly expresses they had a good time and/or appreciates the venue. It includes comments that consider the science centre is not missing anything.	"Don't close it!" "Nice exhibition. I would come again. Thanks". "It was great fun".	"They should add more interactive displays to the rest of the museum" (it does not directly say DW was great, it is a suggestion for the galleries, goes in Other). "Thanks for having a half price day" (the appreciation is not about the venue, but the ticket price, it goes to Other).
Tropical Forest (TF)	Mentions of the Tropical Forest.	"I didn't really like that people want take/touch the butterflies".	"To do more show about butterflys maybe one every hour" (it fits better in Sug).
Suggestions (Sug)	Suggestions of changes in Discovery World.	"Make it more/have more places for teens". "Make the descriptions shorter and more fun to read". "Decrease price". "Please maka a sealife exhibit".	"They should add more interactive displays to the rest of the museum" (it is not a suggestion for the Science Centre, it goes to Other).
Avoiding Comment (AC)	Technically they are comments, but they do not say anything, it was just to fill something in or they don't know what to say.	"I do not know". "Nope".	"None it's perfict" ("None" would have been AC, but the inclusion of "it's perfict" makes it EA).
Other	Any comment that does not fit in any of the above categories.	"Already written." (it is a post-comment, but there is no pre-comment, probably is referring to other open question).	

Coding manual D2: Comments - Tūhura

This code was used in comments from Surveys 2, 3 and 4.

Table

Coding manual D2

Category	Definition	Inclusions	Exclusions
Enjoyment / Appreciation (EA)	Comments where visitors express they had a good time and/or appreciate the venue. It includes comments that consider the science centre is not missing anything.	"It is a fantastic experience thank you". "A great resource for Dunedin". "Keep up the good work".	
Suggestions (Sug)	Suggestions of changes in Tūhura.	"Cheaper for rate payers plz".	"Add air hockey and mind battle" (these are

		"Please make exhibition about earth spinning".	exhibits from Discovery World, it goes to DW). "Fortnite" (it is a game, but it doesn't suggest to be added, it goes to Other).
Miss DW or something from DW (DW)	Mentions to something they miss from Discovery World.	"Why did you close discovery world?" "Bring back the piano".	
Avoiding Comment (AC)	Technically they are comments, but they do not say anything, it was just to fill something in or they don't know what to say.	"Not sure". "No".	"All good" (short answer, but if all is good, it goes to EA).
Other (Ot)	Any comment that does not fit in any of the above categories.	"I have mine on questions board". "You go glen coco".	

Coding manual E: "What do you miss from Discovery World?"

This code is about counting the exhibits visitors missed the most from Discovery World in Survey 2. Technically, each exhibit was its own code. Each exhibit was counted individually, even if due to their low frequency they were places in the category "Other". Example: "The aquarium. Air hockey and games" (2 counts in OE, 1 in NS).

Table
Coding manual E

Category	Definition	Inclusions	Exclusions
Piano (P)	The piano with the big keys con the floor.	"The walk on piano." "Giant piano keys"	
Air Hockey (AH)	Air hockey table.	"The air hockey".	
Mind Ball (MB)	Mind Ball, where participants relaxed so a sensor detected a change in alpha and theta waves to move a ball.	"The mind concentration game". "The focus test."	
Misses something not specific or not an exhibit (NS)	Tick this one off if they miss something, but don't name particular exhibits, or if what they miss is not an exhibit. Do not include planetarium and Tropical Forest.	"Most of the things in discovery world I wish they kept all of them". "Opportunity for kids to concentrate for better understanding". "More things for younger children".	
Don't Miss it (DM)	They don't miss anything, are unsure or can't even remember how DW was.	"No". "Too long to remember".	
Planetarium (Pl)	Apparently, some people think the planetarium is gone.	"Planetarium".	
Other Exhibit (OE)	Exhibits that are not listed above, including the Science Show.	"Air gun with balls and hoops" (Hover Ball).	"Rainfall comparison info was good" (this panel was from

		"The smelly things" (Smell This). "Nothing comes to mind. The hot air balloon was pretty cool." (it was an old exhibit). "I miss that bubble thing" (not sure what exhibit it is, but it is an exhibit). "Massive piano Triangle mirror room" (one is Kaleidoscope, here in OE, and the other is Piano).	Tropical Forest, it goes to Other).
Other (Ot)	Any comment that does not fit in any of the above categories.	"Just enjoyed the lot" (not sure if enjoyed Tūhura or DW). "Probably".	"Everything is pretty similar" (if consider similar, they are not missing anything, it goes to DM).

F.2 Coding manual for interviews

There were 2 interviews, one before and one after the redevelopment. Some of the questions were asked in both interviews (to see if there was consistency in their answers before and after the redevelopment process), but one of them only answered the 1st one and 4 answered both in a single larger interview. The single interview was kept, and the rest pre/posts were combined, so that everybody has 1 long interview (repeated questions' answers were merged).

Most of the codes were linked to individual questions, but if info related to the code appears somewhere else in the interview, it is still taken into account.

The Maori curator and technician interviews were not coded because they were individual and specifically tailored at their duties.

Coding manual F: Characterizing a good science centre

In this code The terms "attraction, retention and expansion (a.k.a. development)" are taken from business models. (+) marks positive inclusions and (-) negative inclusions, but it the end, what will be reported is the positive things to look for, only in the case something negative cannot be transformed into something positive, it will be classified as something to avoid.

Table

Coding manual F

Category	**Subcategories**	**Definition**	**Inclusions**	**Exclusions**
Exhibits Attraction	Visual appeal; Diversity of topics and approaches;	Characteristics that attract people to the exhibit.	"very big showpiece (+)", "it doesn't look nice (-)","I think that a child or whatever they'll just see it and be like 'oh, it's kind of cool' and then move along (-)","there's enough diversity here"	
Exhibits Retention	Ease of use; Interactivity; Enjoyment; Testing diversity;	Characteristics that make visitors stay and engage with the exhibit.	"I think it's a really frustrating exhibit (-)", "I like the interactivity of the Animation Station (-)","On the child-like level, I enjoy shooting people with the Air	"cause it's such a different sense" (Attraction to try something different)

			Gun (-)","I like the way that you can play with it a lot and each time it's different (+)",	
Exhibits Expansion	Phenomena exposure; Scientific explanation (panel information + science learning);	What makes the effect of the exhibit transcend beyond the moment, leaving the visitor with something to take away.	"it has been more of a genuinely good learning tool that I thought it would be (+)", "it's just a really simple, cool (+)", "it does make it more into a playground as opposed to a science centre (-)", "I don't think there is a very good explanation of the science as well (-)","there's literally just a paragraph and some of these interactives are far more complicated than to be explained by one paragraph (-)","we haven't got it going as well as we could do (-)","it's got good scientific concepts (+)"	"seems quite intellectual" (it "seems", it fits in Attraction)
Exhibition Experience	Organization; Coherence, Use of spaces, Aesthetics; Transgenerational; Self-explanation; Inspiration; Demographics enhancement; Use of potential;	Comments related to the experience as a whole, not by single exhibits.	"Discovery Theater, a place where we can do presentations (+)","it's far more adult friendly than Discovery World", "you get families playing in the same thing", "it is only focused on children (-)","evokes conversation around science limitlessly in every way (+)", "I'd like to see some flow (-)", "it's got the depth of the narrative, the storyline (+)"	"I think Discovery World now is more of a playground" (fits better in Exhibits Expansion, as there is a take-away)
Institutional Factors	Running costs; Dimensions; Maintenance; Financial sustainability; Museum's recognition; Age; Maori conversation;	What is looked for by the science centre as an institution.	"very important come through, from all of the world, and they seem to enjoy it (-)","everybody that lives in Dunedin has been to Discovery World at some point (-)"	"the repeat visitors, people who come and then upgrade to annual pass for example" (in the context, it was mentioned as an expected sign of highest engagement, fits better in Expansion), "sometimes it doesn't work" (if the context is the cost of continually repairing the exhibit, it fits here, but in this case it's related to affecting the visitor's experience,

				Exhibition Experience.
Other	Personal tastes	Any other category that is important to acknowledge, but does not fit in the previous ones	"I love music (+)", "personally that's not a thing that I look for in a science centre (-)"	

<u>Coding manual G: Description of the role (DM and SC)</u>

Table

Coding manual G

Category	Definition	Inclusions	Exclusions
Decision making	Making decisions and managing related to the science centre or the redevelopment	"strategic direction of the museum"	
Scientist	Doing research in science	"I'm also a researcher"	
Transmitter of enthusiasm	Helping people to feel engaged with science	"make a warm environment"	
Facilitator of learning	Helping people to learn science	"to teach people about science"	

<u>Coding manual H: Level of interaction with exhibits</u>

This one was coded in magnitude coding.

Table

Coding manual H

Category	Definition	Inclusions	Exclusions
Low	Has interacted with a few of the exhibits		
Medium	Has interacted with most of them, but not all	"I have interacted with most of the exhibits"	
High	Has interacted with all of the exhibits	"I would've interacted with all of them at some point"	

<u>Coding manual I: Level of knowledge of the Dodd-Walls Centre</u>

This one was coded in magnitude coding.

Table

Coding manual I

Category	Definition	Inclusions	Exclusions
Low	From not knowing anything to having a vague idea	"It's, isn't it part of the Centre for Innovation?"	
Medium	They can mention what is the field of work or what is the relation with the museum, without knowing details or having them wrong	"Funnily enough, I think I don't know enough… it's linked to all the quantum physics that's happening; at least that's what I would say, I guess"	
High	They can explain what it is and does without being dubious.	"it's one of the centres for research excellence, funded by the government"	

Coding manual J: Favourite exhibits

This one was coded in magnitude coding.

Table

Coding manual J

Category	Definition	Inclusions	Exclusions
Low appreciation	List the exhibits that are the personally least favourites at the science centre		
High appreciation	List the exhibits that are the personally most favourites at the science centre		

Coding manual K: Definition of science

Table

Coding manual K

Category	Definition	Inclusions	Exclusions
Tool to generate knowledge/understanding	Science as something you use, like a tool, to understand the world (analogy: the hammer in a forge)	"science is just the tools of trying to learn about the world"	
Process of generation of knowledge/understanding	Science as the process to understand the world (analogy: forging)	"Learning about the world" "just the study of the way the world works"	"the thing that you apply to a situation to help you understand it" (Tool)
Scientific method	Specific mentions to "scientific method", "method" or the steps in it, such as formulation of hypothesis	"through systematic approach of trial and testing"	
Questioning	Science as making questions and/or answering questions	"always asking questions"	

Coding manual L: Definition of science outreach / science communication

Both definitions used the same code, although they were coded independently.

Table

Coding manual L

Category	Definition	Inclusions	Exclusions
Bridge between scientists and laypeople	There's a gap between scientists and the public that needs to be bridged by making science accessible	"Being the bridge between academics and incredibly interesting new science and people who want to know it", "just getting scientific concepts and putting them in a manner that is accessible to who you are talking to"	
Teaching science	Facilitating visitors to understand science and scientific concepts	"giving them some skills, tools or excitement for science"	
Sparking excitement for science	Facilitating visitors to feel engaged with science	"giving them some skills, tools or excitement for science"	

Taking science communication to another place	In the case of outreach, doing science communication out of the usual place (e.g., outside the museum).	"science outreach is the people that come to new places with a tent or something like that and teach you some science"	

<u>Coding manual M: Aspects of science to get across.</u>

Table

Coding manual M

Category	Definition	Inclusions	Exclusions
Understanding of what science is	Awareness and understanding of what science is and how it works	"aspects of science are kind of all around us"	
Application	View science as something that can be applied/used	"to teach the application of science"	
Engagement / Curiosity	Start wondering how things work / Engagement with science, so they can continue learning by themselves	"if they came once and they learnt about cool science is"	
Scientific method	Specific mentions to "scientific method", "method" or indirectly through at least two steps.	"making hypotheses and aims", "the idea of observing things and, you know, testing things, and sort of enquiring"	
Inquisitive thinking	To accept theories only after revising them thoughtfully.	"Teaching people how to think" "it's getting the enquiring mind, but also, maybe to some extent, kind of, in a good way, like a skeptical mind"	"a sense of inquiry or curiosity into science in the first place" (Engagement / Curiosity)
Information	Information, facts, concepts	"getting a few key concepts is good"	

F.3 Inter-coder reliability for surveys' open questions

AA stands for the Average pairwise percent of agreement, KA for the Krippendorff's Alpha and PDR for the Positive Decisions Ratio.

Table

Code B1 ("It was cool learning about..." in Discovery World). Sample of 17 of 173 responses (10%)

Code	Example	AA	KA	PDR
Tropical Forest	"How the butterflies are sourced"	100%	1.000	.294
Planetarium	"Southern lights"	100%	1.000	.059
Science Show	"Sound waves"	100%	1.000	.176
Light and Electromagnetism	"Electricity in the neon room"	100%	1.000	.118
Not-Light and Electromagnetism	"Thought waves"	100%	1.000	.353
Science in General	"Heaps of things"	100%	1.000	.235
Other	"I also thought it was cool seeing all the different things you can do"	96%	.000	.020

Table

Code B2 ("It was cool learning about...", "Can you give an example of something you learnt?" and "Have your say! Let us know what you liked or didn't like of Tūhura's exhibits, or any comment you have" in Tūhura). Sample of 49 of 610 responses (8%)

Code	Example	AA	KA	PDR
Tropical Forest	"The butterflies were amazing."	99%	.943	.136
Planetarium		97%	.702	.048
General Light	"Lighting effects"	NA	NA	NA
Light and Electromagnetism	"What three colours make a white light"	95%	.873	.306
Biology	"How your body works"	93%	.845	.320
Psychology, Art and Culture	"The colour stuff and how it efects your moods"	100%	1.000	.136
Earth Sciences	"How tornados look"	99%	.893	.068
Mechanics	"Effect of torque"	99%	.893	.068
Science in General	"How things work"	100%	1.000	.102
Other	"Its was really fun and i think alot of people would like it."	95%	.610	.075

Table

Code C1 ("It would be cool if Discovery World had an exhibit about..." in Tūhura). Sample of 20 of 132 responses (15%)

Code	Example	AA	KA	PDR
Animals	"Ummm mini chickens"	100%	1.000	.100
Extinct Animals	"Dinosaurs"	100%	1.000	.150
Biological Systems	"Memory genetics animal adaptations"	97%	.882	.167
Light and Electromagnetism	"Refraction of light"	100%	1.000	.050
Earth Sciences	"Earthquakes"	97%	.737	.067
Universe	"Space"	97%	.491	.033
Sound	"Music"	100%	1.000	.050
Forces and Mechanics	"how fast you kick or throw a ball"	97%	.737	.067
Technology	"Aviation"	97%	.491	.033
Science in General	"More nz science things"	97%	.491	.033
Good as Is	"Hard to fault what u have already Y"	NA	NA	NA
Avoiding Comment	"I dont know"	100%	1.000	.100
Other	"I would need to research that"	97%	.882	.167

Table

Code C2 ("It would be cool if Tūhura had an exhibit about..."). Sample of 17 of 298 responses (6%)

Code	Example	AA	KA	PDR
Animals	"Different breeds of fish and birds"	96%	.878	.196
Extinct Animals	"Dinosaurs"	100%	1.000	.118
Biological Systems	"The human brain"	100%	1.000	.059
Physics and Chemistry	"Atoms"	100%	1.000	.118
Earth Sciences	"athmosphere"	96%	.886	.216
Universe	"Space"	100%	1.000	.118
Transport and Technology	"Something about planes or streamlining cars / vehicles"	100%	1.000	.118
Science in General	"Anything is good"	NA	NA	NA
Good as Is	"I enjoy everything here as an adult not sure what else id like to see"	NA	NA	NA
Avoiding Comment	"Not sure"	92%	.306	.059
Other	"Tacos"	96%	.783	.098

Table

Code D1 (Discovery World's Comments). Sample of 16 of 40 responses (40%)

Code	Example	AA	KA	PDR
Enjoyment / Appreciation	"Nice exhibition. I would come again. Thanks"	96%	.918	.479
Tropical Forest	"i loved seeing the different butterflys"	100%	1.000	.125
Suggestions	"Make it more/have more places for teens"	92%	.832	.417
Avoiding Comment	"Nope"	96%	.000	.021
Other	"Thanks for having a half price day"	88%	.440	.104

Table

Code D2 (Tūhura's Comments). Sample of 18 of 220 responses (8%)

Code	Example	AA	KA	PDR
Enjoyment/ Appreciation	"Awesome wish we could have stayed longer"	100%	1.000	.611
Suggestions	"More explanations"	96%	.839	.130
Miss Discovery World	"Please bring back the mind control"	100%	1.000	.056
Avoiding Comment	"No"	100%	1.000	.222
Other	"Fun to see if the app works"	NA	NA	NA

Table

Code E ("What do you miss from Discovery World?" in Tūhura). Sample of 18 of 149 responses (12%)

Code	Example	AA	KA	PDR
Piano	"Giant piano"	96%	.490	.037
Air Hockey	"The air hockey"	100%	1.000	.111
Mind Ball	"The brainwave thing"	100%	1.000	.111
Not Specific	"Some exhibits"	96%	.735	.074
Don't Miss It	"Not really"	96%	.922	.370
Other Exhibit	"The smelly things"	96%	.901	.241
Other	"Opportunity for kids to concentrate for better understanding"	89%	.204	.074

Appendix G. Extract of the interview with Dr. Oleg Abramov

Interviewer: How do you become a space scientist? [question asked by Fergus]

Dr. Abramov: So, for me, this was a long road that started when I was, probably around 8 years old. I was in third grade and, I just started to get a little bit interested about space, but of course as a child I was also interested in many other things. So, I went to the library, the school library and I asked for a couple of books about space. I didn´t really know specific details of what I was interested in, and the two books that the librarian gave me were both on other planets in our solar system. As I started reading them, I discovered that we were already sending spacecraft to land on these planets to take samples of the atmosphere and the ground and take pictures and that human missions to these planets in our solar system would… I would probably live to see them, and I found that very, very exciting, that I lived in a time when human beings are about to go to other planets and that inspired me to learn more about other planets in our solar system. Initially I thought my main interest was astronomy and I wrote an essay at school, when I was in 4th grade about why I wanted to be an astronomer, but I didn´t know there was such a thing as a planetary scientist or space scientist at the time, but in my essay all I wrote about was planets. So, I kept that interest and I started learning more and more and more about planets in our solar system and other objects, comets, asteroids, what kind of spacecraft we're sending out there and what kind of instruments they had and I was very hoping to find out and what the outstanding questions are. Basically, what the mysteries are of the solar system. So, that inspired me to find out where I could get a graduate degree in planetary sciences and one of the best places in the world, I found, was at the University of Arizona, so I applied there, actually several times before I was accepted. It was a difficult journey, but it was something I wanted to do, so I kept studying more and taking more exams and reapplying. On my fourth try to get into the program I was finally accepted I got my PhD at the University of Arizona in Planetary Sciences and that allowed me a lot of opportunities to work in new places like the Lunar and Planetary Institute, in Houston, places that are specifically specialized in the study of other planets and analysing, for example, samples [unintelligible] from the moon or meteorites that came from other planets. So, now I'm working for an institute called the Planetary Sciences Institute, which is headquarted in Tucson, Arizona, but I'm able to be here in New Zealand, because I'm able to work remotely and the work that I do is mostly computer modelling that I can do from anywhere in the world. So, this has opened a lot of exciting possibilities for me. [I: That's an excellent advise, don't ever give up] Yeah, don't give up, there will be challenges all over the way, some things will not work out and you have to persevere and follow your dreams, follow what you are really interested in. That's very important.

Interviewer: Is Mars big enough for all humans to live on? [question asked by Fergus]
Dr. Abramov: For all humans? Right, so, all humans that are on Earth, could they, all of them, live in Mars? That's an interesting question. Right now, our technology is such we wouldn't be able to support that many humans living on Mars, because we would need places for them to live, we'd need life support equipment. Mars is, unlike Earth, has an atmosphere of Carbon Dioxide, that's about 1% of ours Earth's density, so our people will have to live in habitats, in domes, and that will take quite a bit of time to build up the infrastructure, to be able to house a lot of people on Mars. Eventually, however, the plan is to make Mars more Earth-like, to thicken up its atmosphere, to plant algae, and eventually plants there. That would enrich the atmosphere in Oxygen. Mars does have a very large amount of water ice in its polar icecaps and in the sub-surface and that water ice can basically be turned into water vapour, it can also be split into Hydrogen and Oxygen. So, using these processes it is possible to densify the atmosphere, make Mars warmer, make Mars where human beings would be able to live much like they do on Earth, and at that point, Mars could potentially sustain the current population of Earth, but that would be probably a couple of hundred years in the future, I imagine.

Interviewer: Are there more planets that can become habitable?* [question asked by Anna]
Dr. Abramov: Alright, yeah, so, are there any planets that we can make habitable? Yes. So, Mars is the most logical place for humans to go, initially to settle, because it's more Earth-like. Other than Mars, there are several other possibilities, so, the Moon, just by virtual, its very close proximity to Earth could eventually be terraformed in much the same way, but it would be a more difficult process because the Moon doesn't have very much in the way of water ice or other we call volatiles elements and compounds that can be made into atmosphere gases. So, it would require, basically, a lot of energy to, for example, liberate Oxygen from the rocks on the Moon and make that Oxygen into atmospheric gas or, transport vast amounts of materials from the Earth to the Moon to make it more Earth-like, but because it's so close to the Earth it's also not impossible to just, it's further in the future. Venus is another potential candidate, but Venus has the opposite problem. Mars has an atmosphere

that’s very rarefied, Venus has a very thick atmosphere and it has a massive greenhouse effect and very high surface temperatures, so, on Venus we’d have to figure out how to make the atmosphere a lot thinner, how to remove that Carbon Dioxide from the atmosphere, which is what we are trying to figure how to do on Earth right now, to combat global warming. So, technologies might be developed that will allow us to do something similar in Venus, to make the atmosphere thinner and make Venus more sustainable for humans to life, more habitable for humans. So, those are probably the main options. There are also moons of Jupiter and Saturn, like Europa and Titan. They are remote, they are very cold, but they do have a lot of water ice and do they do have other resources that could potentially be useful for humans, but life on those locations would be quite different than life that we are used to on Earth. So, I think Mars, followed by the Moon, followed by Venus are probable the most likely options of places that we can make habitable for humans in the future.